

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 20:10:33 ; Search time 38.03 Seconds

(without alignments)
1274.701 Million cell updates/sec

Title: US-09-938-330-20

Perfect score: 6954
Sequence: 1 MAPRLALSYLLPLHCAICT.....PTGPTLCVLPDQSLRGHT 1252

Indexing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Matched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	3614	52.0	1201	AT53_HUMAN	O15072 homo sapien
2	3326	47.8	1205	AT52_BOVIN	P29331 b adams-2
3	3299.5	47.4	1211	AT52_HUMAN	O95450 h adams-2
4	1304	18.8	1077	AT10_HUMAN	O91324 homo sapien
5	1263.5	18.2	1629	AT59_HUMAN	O952n4 homo sapien
6	1244	17.9	1533	AT12_HUMAN	P58397 homo sapien
7	1119.5	16.1	967	AT51_RAT	O9wq91 rattus norv
8	1108.5	15.9	968	AT51_MOUSE	P97857 mus musculu
9	1098.5	15.8	967	AT51_HUMAN	O9uh18 homo sapien
10	1068	15.4	930	AT55_HUMAN	O9una0 homo sapien
11	1062.5	15.3	997	AT57_HUMAN	O9ukp4 homo sapien
12	1040.5	15.0	860	AT56_HUMAN	O9ukp5 homo sapien
13	1027	14.8	930	AT55_MOUSE	O91001 mus musculu
14	1013.5	14.6	890	AT58_HUMAN	O9up29 homo sapien
15	945	13.6	905	AT58_MOUSE	P57110 mus musculu
16	898.5	12.9	837	AT54_HUMAN	O75173 homo sapien
17	807	11.6	630	AT54_RAT	O9esap7 rattus norv
18	528	7.6	450	AT10_MOUSE	P58459 mus musculu
19	417.5	6.0	826	AD08_MOUSE	O05910 mus musculu
20	398.5	5.7	571	DIS1_BOVIA	P30431 boithops ja
21	393	5.7	824	AD08_HUMAN	P78325 homo sapien
22	380.5	5.5	207	AT55_BOVIN	O9ct92 bos sapien
23	346	5.0	956	AD19_HUMAN	O9uh03 homo sapien
24	341	4.9	920	AD19_MOUSE	O35674 mus musculu
25	328	4.7	774	AD28_MOUSE	O91166 mus musculu
26	319	4.6	245	AT54_BOVIN	O9ct93 bos taurus
27	311.5	4.5	776	AD28_MACPA	O9xsl6 macaca fasc
28	309	4.4	788	AD07_MOUSE	O35227 mus musculu
29	306.5	4.4	776	AD21_MACPA	O28475 macaca fasc
30	306	4.4	722	AD21_HUMAN	O9ukf8 homo sapien
31	302	4.3	790	AD30_HUMAN	O9ukf2 homo sapien
32	297.5	4.3	760	AD25_MOUSE	O9rl59 mus musculu
33	295	4.2	1077	SM5A_MOUSE	O62217 mus musculu

Description

Cite list
No 1

34	294.5	4.2	814	AD15_HUMAN	O13444 homo sapien
35	294	4.2	754	AD07_HUMAN	O9h2u9 homo sapien
36	294	4.2	775	AD28_HUMAN	O9uk92 homo sapien
37	294	4.2	1074	SM5A_HUMAN	O13591 homo sapien
38	291.5	4.2	1093	SM5B_MOUSE	O60519 mus musculu
39	286	4.1	816	AD15_RAT	O9qyv0 r adam 15 p
40	276.5	4.0	729	AD21_MOUSE	O9j176 mus musculu
41	276.5	4.0	1584	BA11_HUMAN	O14514 homo sapien
42	275	4.0	815	AD15_MOUSE	O88839 mus musculu
43	274.5	3.9	726	AD20_HUMAN	O43506 homo sapien
44	273	3.9	909	AD12_HUMAN	O43184 homo sapien
45	272	3.9	903	AD12_MOUSE	O61824 mus musculu

ALIGNMENTS

RESULT 1
ID AT53_HUMAN STANDARD: PRT: 1201 AA.
AC O15072;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (Fragment).
GN ADAMTS3 OR KIAA0366.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,
RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."
RT DNA Res. 4:141-150(1997).
CC -1- COFACTOR: BINDS TIZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC EMBL: AB002364; BAA20821.1;
CC MEROPS: M12.220;
CC MIM: 605011;
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_MTPepidse.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF00090; TSP_1; 4.
DR SMART: SM00209; TSP1; 4.

FT VARSPLIC 567 1211 MISSING (IN ISOFORM SPNP1).
 SQ SEQUENCE 1211 AA; 134722 MW; BECEP25C3CAD2D CRC64;

Query Match 47.4%; Score 3299.5; DB 1; Length 1211;
 Best Local Similarity 52.6%; Pred. No. 2,3e-164;
 Matches 635; Conservative 156; Mismatches 353; Indels 63; Gaps 20;

1 MAPRLALSYLT-----PLHCACTAAGSTPPLHLHSGKLSIDYGVTPCSTDFRG 50
 12 LCPALLLLLLLPPPLPPPPANARLAADPPGGL---GGAEKILLAVPRTDAG 68
 51 RELSHVSGPAAASAGSWVDTPPLPRSSHLRVARSPPLRGGLMPGRGRSLTENY 110
 69 RLVSHVSA-ATSRAGVARRAP-----VRPSPGNG--EPEEGSH-LFVNV 113
 111 TVFGKELHLRPNRLVYVPGSSVEMQDFRELFRQLRQCVYTGVTGMPGA-AVAIS 169
 114 TVFGKDLHLRPNRLVAPATIMQGEKGTTRVEPLGLSCLYGVDALEASSVALS 173
 170 NCDGLAGLIRDTSTDFTEPLERGQOEKA-SGRTHVYVREAVQOMAEPDG-DLINEA 227
 174 NODGLAGLIRDEEFTLEPLEKGLAQAQVAVYRRPPTSPPLGGPQALDTGASL 233
 228 FELGDLPLNLGLVGDQIDTERK-RRNAKPGSYSIEVLVVDSDVAFHGEKHYONYLT 286
 234 DSDLSRALGVLEEHANSSRRARRHADDDVIEVLVGVDSVYGFHKEHYOKYLT 293
 287 LMNIYDELTHHESIGVHNTALVYLWVGRKQSLILEKPNPSRLSEQVCRMAHSGOQROD 346
 294 LMNIYDELTHHESIGVHNTALVYLWVGRKQSLILEKPNPSRLSEQVCRMAHSGOQROD 353
 347 PSHAEHNVHVLTRDQFGPS---GYAPVTGMCPLRSCALNHEDGSSAFVLAHETGV 403
 354 TGHDEYHDHAIJFLRQDGGPGMGVAPYVGMCHPVASCTLNHEDGSSAFVLAHETGV 413
 404 LGMEHDGOGNCADETSIGSMAPLVQAAHRRHWSKSLERLPSYDCLLDDEPDP 463
 414 LGMEHDGOGNCADETSIGSMAPLVQAAHRRHWSKSLERLPSYDCLLDDEPDP 473
 464 AMPPELPGIYNSMDEQCFDFSGYQTCALFRTPEPCQOLMCHSPNDPCKTKKGP 523
 474 DMPALPOLPGHYSMNOCRFDFGLCMCTARTEDPCKQOLMCHSPNDPCKTKKGP 533
 524 LDGECAPKMKCFKGRKHTKMSPEQTYGQDGSWMTKFGSCNSGCGVRSRSCNNPS 583
 534 LDGTCAPKMKCFKGRKHTKMSPEQTYGQDGSWMTKFGSCNSGCGVRSRSCNNPS 592
 584 PAYGRCPLGPMFEYVYCNSECPGYEDFEDRAOCCAKRNSYVYHONAKHSWYPERDDA 643
 593 PANGRCPLGPMFEYVYCNSECPGYEDFEDRAOCCAKRNSYVYHONAKHSWYPERDDA 652
 644 QKCELICOSADTDVYFMNVHDGTRGSYRDEYSVARGECVPCDKEVSMKADKC 703
 653 ERCHLYESETREGEVYSMKRMVHDGTRGSYRDEYSVARGECVPCDKEVSMKADKC 712
 704 GVGCGDSHRTYKGLTGKASKQAGALKLVQIPAGARHIQLEALEKSPHSRYVKNQVTS 763
 713 GVGCGDSHRTYKGLTGKASKQAGALKLVQIPAGARHIQLEALEKSPHSRYVKNQVTS 772
 764 FILNPKG-EATSRFTAMLEWEDAVEDAKSLKTSGLPEALILMLPRDEGPRSL 822
 773 FILNENDVYASSKFTLAVGEVYEDDEGRRLTQMGILHGTITVVLVY-VGTRKSL 830
 823 AKKYVHEDLLPLIGSNVYLEMDTYEALKSWAPCSKACGGGIGFTKYGCRRRRDHNM 882
 831 TYKTMHEDSLN-VDDNNVLEDSVYEWALKKWSKPCGSGGSGFTKYGCRRRRDHNM 889
 883 VORHLCDHKRRPKPIRRRCNOHPCSPVYVTEBWKACSGCKGLGVORGLQCLPLSNG 942
 890 VHRGFCALSKRAIRACNPQSCSPVYVTEBWKACSGCKGLGVORGLQCLPLSNG 949
 943 THKVPAAKACAGDRPARPCLRVPCPAQWRLGAMSQCSATCGEGIQORVVCRTANSL 1002

Db 950 TTRSVHAKHCNDARBPESRRACSRLELCPGRMAGPMSQSVTCGNGTDERPACTADDSF 1009
 1003 GHOEDRPPTVQVCSLPACGCHONSIVRADYWEELGTREGQVQSGPLHPINKISMCA 1062
 1010 GICQERPETATRCRQPCPRNISDPSSKSYV-----VQMLSRPPDPSIRKISS--- 1059
 1063 ABPCTGDRSVFCOMBLVDLRYSIPGYHRLCCVSC-----IKKASGPNPGP-----DPG 1110
 1060 KHCQGDGKSIKCFRMVYLSRYSIPGYHRLCCVSC-----IKKASGPNPGP-----DPG 1119
 1111 PSLPPEPSTPGSPLDPGDDPADAEPPKPTGSSDHQGRATOLPGALDTSPTQHP-F 1169
 1120 MPTLVPVYVAMEVRPSPSTPLEVPLMNSTWATEDHETNAVDEPKIHGLEDEVQPNL 1179
 1170 APETPIP 1176
 Db 1180 IPRRSP 1186

RESULT 4
 AT10_HUMAN STANDARD; PRT; 1077 AA.
 ID 09H324:
 AC 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
 GN ADAMTS10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RA APle S.S.;
 RP SEQUENCE FROM N.A.
 RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple thrombospondin type 1 repeats."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF163762; AAC35563.1;
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSP1.
 CC InterPro: IPR000130; zn_MTPepidase.
 CC Pfam: PF01562; Pep_M12B_propep. 1.
 CC Pfam: PF01421; Reprolysin. 1.
 CC Pfam: PF00090; TSP_1. 5.
 CC SMART: SM00209; TSP1. 5.
 CC PROSITE: PS50215; ADAM_MEPRO. 1.
 CC PROSITE: PS50092; TSP1. 2.
 CC PROSITE: PS00142; ZINC_PROTEASE. 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FAUSE_NEG.
 CC Hydrolase: Metalloprotease; Zinc; Glycoprotein; Zymogen;
 CC Repeat; Extracellular matrix.

```

FT NON_TER 1 1
FT PROPER <1 207
FT CHAIN 208 1077
FT METAL 366 366
FT ACT SITE 367 367
FT METAL 370 370
FT METAL 376 376
FT METAL 376 376
FT DOMAIN 434 520
FT DOMAIN 578 679
FT DOMAIN 680 802
FT DOMAIN 521 577
FT DOMAIN 799 860
FT DOMAIN 862 918
FT DOMAIN 922 976
FT DOMAIN 981 1031
FT CARBOHYD 64 64
FT CARBOHYD 196 196
FT CARBOHYD 297 297
FT CARBOHYD 714 714
FT CARBOHYD 769 769
FT CARBOHYD 866 866
SEQUENCE 1077 AA; 118072 MW; 391ADE18DCBFB587 CRC64;

```

Query Match 18.8%; Score 1304; DB 1; Length 1077;
 Best Local Similarity 31.0%; Pred. No. 8.9e-61;
 Matches 359; Conservative 144; Mismatches 484; Indels 172; Gaps 44;

```

25 RTELHLSGKLDYGVTPCSTDFRGRFLSHVYSGPAASAGSVVDPPTLRHSSLR 84
1 RSODEFLS-SLESEYELAFPTFVDHNGALLAF-----SPP-PRORR-- 40
85 VARSPLHGGTLNMGVRGRHSLYFNVYVFGKELRLRPRRLVGVSSWEMEDREL 144
41 -----GT-----GATAESRLFYKVASPSTHFLNLTSSRLAGHVSVEYTRGLAW 88
145 ROPLEOCVYTGVTGVMGGA-VAISNCDGLAGLIRTSDFLEPLE--RGOQEKAS 200
89 QRAARPHLYAGHLQOGASSHVALSTCGHLGLIVADEDEYLEPLHGPKRSRPEES 148
201 GTHVYVREAVQGEWMAEPDGLINEARGLD--LPNLGLYGDOLG-DTEKRRHAP 256
149 G-PHYVYKRSLLRHPHLDTACGVDEKPKWKPMMRTKLPPRAPRLGNTER--GDP 203
257 G-----TEVLLYVDSVVRPHGKENVYVLTLMIVDEYDESLGHINIALY 309
204 GKRSVGSRRRYVEYLVADKMAVAYHGRRDYEQYVLAIMIVAKLFDDSLGSTVNI 263
310 RLIMVGYRSLIERGNPSRLDQVCRMAHSQGRDPSAER-----HDHYVF 358
264 RLILLLEDDQPLEITH-HAGKSLDSFCW--OKSIVNHSGHNAIPENGVAHDTAVL 318
359 LRDPF-----GRPGYAVPTGKCHLRSALNHDEGSSAVVIAHENGVLGMHDG 410
319 ITRYDICIYKKNPCSTGLIAPRGKRCERSCSVNEDGLPQATTIAHETGHTFGMHDG 378
411 OGNGC-ADETSLSGVMAFLVQAARFHFMSRCSKLESLRYLP--YDCLLDPEPAPQ 467
379 VGNSCCARQODPAKIMAAHITKTPFVWSSCNDRYITSLDGLGLCLNNRPRDQFVY 438
468 PRLPGINSMDQCRFDGSGYQICLAFRRFPCKQKLMCHPNPFFCKTKKPPRDP 527
439 PTVAPQADADADOCRFQGVKSRQC--KGEVCSSELMCLSKSNR--CTNISPAAEGT 493
528 EAPARQ---WCFKGHCT--WKSPPQYGGDGMSSWTKFSGSCSGGVRSHRSRSCNN 581
494 LQDHTIDKGCYKRYCVYFGSRPE---GVGDAMPPTPMDCCRTCGGGSRRSCDS 550
582 PSFAYGPRCLPMFEYGVNSECPTYEEDFRAQOCAR-----RNSYVHQAQKSWY 635
551 PRTTIGKCYCLGERRRHRSCNTDPCPSQDFREYQCSSEFISIPRKFY-----KWK 603
636 PTEPDDAOKCELICOSADTGVYFMNVYHDTGTCGRDYYSVCARCEGVPCGDKMEVG 695

```

```

Db 604 TYR-GGGVACSLTSLAEFNFYTERAAAVYDGTCCR-PDVTIDCVSGECNVGCDRLG 661
Qy 696 SMADKRCVCGGDNDHCTVAGTGLKASKOGALKVQIPAGARHIQIEALESPPRSV 755
Db 662 SDIREKRCVCGGDGACETIEGVSPASPGAGYEDVWIRKGSVHITQDNLNLSHIA 721
Qy 756 VKNQVGSFLLNPKGKEATSRFTAMGLEWEDAVDAKESLTKSGPLDEAIIALLPTE 815
Db 722 LKQDQESLLEGLRPTQPHRLPLAGTTFQLRQPDQVQSLALPINASLIVMLATE 781
Qy 816 GGRSSIAKY--VIEDLLPLIGSNVLLIEMDTYEWALKSMAPCSACGGGQIFRYG 873
Db 782 ---LPALRYFNAPRIARDLPL-----PXSMTYAFWTGCSAQCGAGGSOVQAVE 825
Qy 874 CRRRRHHWQRLCD-HKKRPPIRRRCNPPCSQPVWTEWGCASRSGCLGVQTRG 932
Db 826 CRNOLSSAAYAPHYCSAHSKLPKR-ORACNTERC-PPDMVGNMNLGCRSC-DAGVRSR 882
Qy 933 IOCLPLSLNCTHRYMPAKACAGDRPARPPCLRVPCPAQWRLGAMSQCSATCGEIQOR 992
Db 883 VVCGRRVSAAEKALDSDACQPRPPVLEACHGPTCCPMAALDMSCTPPSCGPIRRHV 942
Qy 993 VVCRTNANSL-----GRCF-GDRPDYQVCSL-----PAX-----CGGNHONSTV 1030
Db 943 VLCKSADHRAITLPPRACHSPAKPPATMRCMLRRCPPARVAVAGEWECASQCGVGRORSV 1002
Qy 1031 RADVWEIGTEPGQWVPQ-SGLPHPIKISSMAAE--PCTGD-----RSVFCOME 1077
Db 1003 RCT-----SHTGQASHCTALRP--PTTQCCAKAKDSTPPGDPRECKDVKAVCYCLV 1055
Qy 1078 VLDRYCSIPGYHRLCYSC 1096
Db 1056 LKFOQCSRAVAFROMCKCTC 1074

```

RESULT 5
 ID: HUMAN STANDARD; PRT; 1629 AA.
 AC O9P2N4; O9NR29;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
 GN ADAMTS9 OR KIAA1312.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC TISSUE-Fetal;
 RX MEDLINE=20396138; PubMed=10936055;
 RA Clark M.E., Kehler G.S., Turbeville L.A., Boyer A., Arden K.A., Maki R.A.;
 RT "ADAMTS 9, a novel member of the ADAM-TS/Metalloprotein gene family";
 RT Genomics 67:343-350(2000).
 RN [2]
 RP SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
 RC TISSUE-Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro";
 RT DNA Res. 7:65-73(2000).
 CC -1 CORFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1 SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.

```

FT CONFLICT 367 367 F -> L (IN REF. 1)
SQ SEQUENCE 1629 AA: 182649 MW: C1C4CEFF58B89A1F CRC64;

Query Match
Best Local Similarity 18.2%; Score 1263.5; DB 1; Length 1629;
Matches 357; Conservative 171; Mismatches 452; Indels 297; Gaps

QY 21 AAGSRPELH-----LSGLSDYGVYVPCSTDFRGFFLSHVSGPAAASAGSMVVDPTPT 75
DB 26 AAAYRKDRLLHPROVKLLLETLLGEYEIYSPRYNALG-----EPEPT 65
QY 76 LPHSSHILRYARSPLLPGCTLMPGRVGRHS-----LYENYVFGKELHLRLRRRLIYV 129
DB 66 ----NHAFKTRTSINSADDPMWPAFASSSSSSSSSQAHRLSAFQOQLPNTANAGTIA 121
QY 130 PGSSVE-----WOEDFRELPROPLROEYVYTGVTGMPGAVAISNCDIAGL 177
DB 122 PLFTVYLLGTPGVNQTKFYSSEEAEL-----KHCFYGYVYNTSEHTAVISLCSGMLGT 175
QY 178 IRTDSTDFTEPLER--GOGEKASGRTHVYVREAYOGEAEPERGDL----- 224
DB 176 FRSHDDYFIEPLQSDHDEDEEONKHIIIRRSAPR--EPSTGHACDTSBKRNH 232
QY 225 -----NEAFGLGDLPLNLGLV-----GDOLGDTERRKRAHPGSY---- 259
DB 233 SKDKKTKRAKNGEIRINTLADVAALNSGLATEAFSAYNKNTDNTREKTRHRTKFLSYR 292
QY 260 -SIEVLVYDDSVYRHEGHEVYONVLTMTVDSIYDESLGYNINATVRLIMVGRQ 318
DB 293 RPEVFLVYADNKNKSYHG-ENLOHILTLMSIVASTYKDPSTGLNLTIVNLIVHNEO 351
QY 319 ---SLSLIERGNPSRSLSEYCVRMASHSOORODPASHAHHDVYFLTRDF-----GPSG 368
DB 352 DGEISIF---NAQTTLKFCQMOHS--KNSPG-GIHMDTAVLLTRDICIAPHDKCOTLG 404
QY 369 YAPVTMCPLRSCALNHEDGSSAFVYTAHEGHVHLGHEHOGQNGCADE--TSLGYMA 426
DB 405 LMEIGTICPYSCSISDEDSGLSTFTTALHGLGFNNPHD-DNNCKKEBVSVPQWMA 463
QY 427 PLVQAAHFHFHWSKSKLESLRYLPS-YDCLLDPPDPAMPQPELPGIYMSDEGCRF 484
DB 464 PRLNFYTNEMMWSKSRKYITEFLDTGEGCLLNEPESRPYPLPVQGLIYVNNKCEL 523
QY 485 DSGSGYQICLARTREPCOKLMCSHPDNPY-FCKTKKRPJLDTGECAFGWCKRGHCITWK 543
DB 524 ITGPGSQV---PYMGQCRILMCNNVNVNHNHGCRQTQHPMADGTCEBPGRHCKYGFVCPK 580
QY 544 SPEQTYQDGGWSSWTKFGSCSGCGGVSRSRSRSCNNPSPAYGRCGLCPMEFYVCNS 603
DB 581 EMDVVPY-ITDSGWSGSPGTCSCRTCGGGIKTAIECKRPREKKNKGKCYVRRKRFKSCNT 639
QY 604 EECPCGTIEDYFAQOCACAKNNSTYVHONAKH-----SNVP-YEPDDDAQCELLCOS 652
DB 640 EECPLCKRDPDEQCA-----HFDGKHFNINGLLPLNVWVWPYKSGILMKDKRCFLCFRY 692
QY 653 ADTGDVYPMNDOVHHDGTCRSYRDYPSVCARECPVPCDCEYVSMADKCGVCGGDSNH 712
DB 693 AGNTAYVYDRLBRVIDGTGCG-QDTINDICVGLCQACQADHYLNSKARRDKCGVGGDSNS 751
QY 713 CRYVAGTLGKASKQAGALKLVDPAGARHIOIE---ALEKSPHRSVYVKNQVYGSFILNP 768
DB 752 CKTVAGTNP--TVHYGVNTVVRIPAGATINIDVRHSHSGTDDDNLTALSSSGEFLNG 809
QY 769 GKKEATSTFTAMG---LEWEDAVEADKESILKTSRPLREAIATIALPLPDGGRSLAK 825
DB 810 NFVYTMARKERIRIGNAVVEYSGS-ETAVERINSTDRIDEJLLOVL---SVGLYNPVVR 865
QY 826 VYIHEDLLPLIGSNVNLLEEM-DTYEM-ALKSAPKSKACGGGIDQFTTKYGCRRRRDHNHY 883
DB 866 Y-----SNINIFEDKPOGFYWNHSGPQWACSKPQGE-RKRKLVTGRRESQQLVY 913
884 ORHLCIDHKKRPKPLRRP-----

```

```

Cc      914 SDORCRLLPQGHITTEPCGTDICDLRWNVASRSECSAQCGLGYFTLIDYAKTSLRDLGKTE 973
Cc      901 -----CNOHP-----CSOPVWVTEWAGSRSCKGLGVOYTRIGIOLLPLSNGT 943
Cc      974 KYDDGFCSSHPKRSNKEKSGEENTGWRASNTAECSKSDG- GTRRRRAICV-----NTR 1028
Cc      944 HXVAPAKACAGDREARPCPLRVCPAOMRLGAMSOCSATCGEIGIOQROYVCFNANSIG 1003
Cc      1029 NDVLDLDSKCTHOKRVITLQRCSEFPCC- QMKSGDMSGLCYCGHGRHROWECQFGEIDLIN 1087
Cc      1004 H--CEGD-REDTVOYCSLPCAGGNHONSTVRADVWELGPEGOWVQSGPLHPINKISSM 1060
Cc      1088 DRMCDPEPTKFTSMOTCOOPPCAS-----WQAG-PMGQCSVTCGGGYL----- 1129
Cc      1061 CAARECTGDRSVFC-----QMEVLD-----RCSIDGYHRLCCVSCIKRASGP 1103
Cc      1130 -----RAVKCIIGTVMSVVDNDNCNATRPDTQDELPSCH----- 1166
Cc      1104 NPGDPPTSLPPSTP 1120
Cc      1167 -PPAPAEPTKSTYSAP 1182

RESULT 6
AT12_HUMAN STANDARD; PRT; 1593 AA.
ID P58397;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE ADAMTS-12 precursor (BC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RX CAL S.; Argueelles J.M.; Fernandez P.L.; Lopez-Otin C.;
RX MEDLINE=21264577; PubMed=11279086;
RT "Identification, characterization, and intracellular processing of
RT ADAMTS-12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats."
RT J. Biol. Chem. 276:17932-17940(2001).
CC -1 COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1 SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -1 TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1 DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1 PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1 PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1 SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1 SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
Cc      This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc      the European Bioinformatics Institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

Cc      or send an email to license@isb-sib.ch.
Cc      -----
Cc      EMBL, AJ250725; CAC20419.1; -.
Cc      DR MIM; 606184; -.
Cc      DR PROSITE; PS02015; ADAM_MEPRO; 1.
Cc      DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
Cc      DR PROSITE; PS00092; TSP1; 2.
Cc      DR PROSITE; PS00142; ZINC_PROTEASE; 1.
Cc      KM Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;
Cc      Repeat; Extracellular matrix.
Cc      FT SIGNAL; 1 25
Cc      FT PROPEP; 26 240
Cc      FT CHAIN; 241 1593
Cc      FT DOMAIN; 465 544
Cc      FT DOMAIN; 545 596
Cc      FT DOMAIN; 597 700
Cc      FT DOMAIN; 701 826
Cc      FT DOMAIN; 827 881
Cc      FT DOMAIN; 886 943
Cc      FT DOMAIN; 947 995
Cc      FT DOMAIN; 996 1315
Cc      FT DOMAIN; 1316 1364
Cc      FT DOMAIN; 1367 1423
Cc      FT DOMAIN; 1426 1471
Cc      FT DOMAIN; 1471 302
Cc      FT DOMAIN; 302 305
Cc      FT SITE; 208 208
Cc      FT METAL; 392 392
Cc      FT ACT_SITE; 393 393
Cc      FT METAL; 396 396
Cc      FT METAL; 402 402
Cc      FT CARBOHYD; 105 105
Cc      FT CARBOHYD; 125 125
Cc      FT CARBOHYD; 215 215
Cc      FT CARBOHYD; 485 485
Cc      FT CARBOHYD; 685 685
Cc      FT CARBOHYD; 790 790
Cc      FT CARBOHYD; 951 951
Cc      FT CARBOHYD; 1104 1104
Cc      FT CARBOHYD; 1275 1275
Cc      FT CARBOHYD; 1300 1300
Cc      FT CARBOHYD; 1320 1320
Cc      FT CARBOHYD; 1371 1371
Cc      FT CARBOHYD; 1378 1378
Cc      FT CARBOHYD; 1503 1503
Cc      SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48B63BD83A3 CRC64;

Query Match 17.9%; Score 1244; DB 1; Length 1593;
Best Local Similarity 28.9%; Pred. No. 1.7e-57;
Matches 383; Conservative 185; Mismatches 554; Indels 202; Gaps 56;

Cc      30 HLSGKLSDYGVTVPCSTDFGRFLSHVSGPAASAGSNVDTPLPHSSHLLVARSF 89
Cc      44 HFIKGLPEYHVVPVRVDASGHLSTGLHYPIYSSRRKDDL----- 85
Cc      90 LHHGGLMPERVARHSLYEVWVYFGKEHLRLPRNRLVVGSSVWQDFRELPROPLR 149
Cc      86 ---GSEDW-----VYRISHEEKDLFNLTVNGFL--SNSYIMKRYGNLSVKKM 132
Cc      150 QE-----CVYTGCV--TGMPCAAVAIISNCDGLAFLPTDSTDFEIEPLERGOOEKASGRT 203
Cc      133 ASSAPLCHLSGLVLAQGTGRTGTAALSCHGLTGFOLPHGDFIEPVKK-HPLVEGYPH 191
Cc      204 HVIYRRAVOOEAEAPDGLAHNAFGLGDLPNLLGLVGQDGLGTEKKR--RAAKPG----- 257
Cc      192 HVIYRRQKV-PEYKEP-----TCGLKDSVNI-----SOKOLWEKKERHNLPSRSIS 238
Cc      258 --SVS-----LEVLVYDSDSVYRFGKEHYOVYVLTIMNIYDEIYHDSLGVINIALVRL 311
Cc      239 RRSISKRWETLVVADTKIETVHGSINVSYILITIMNVTGFLNPSISGMALHIYVRL 298
Cc      312 IMV-GYROSLSLIERGNPSRSLDQVCWMAHSQORODPSHAHHHDHVVFLTRDQFGS--- 367

```

```

Db 299 ILLEEEQGLKIVH--HAERTLSSFCCKMOKSINPKSDLNLPNHHVAVLITTKDKIACAGNR 356
QY 368 -----GYAVYTGKCHPLRSCALNHEDGSSAFVIAHETGHLVGMEDHOGNGCAGDETSLG 422
Db 357 PETTLTISHSGMCPHRSKNINENSGPLAFTIAHLEHSGFIQHOHCKENDCEPVGRHP 416
QY 423 SVAPALVQAFFHFRHMSRCKSLSLRYLPSVD-----CLDDPDPAMPPOPELPGINYS 477
Db 417 YMSRLOVDPTPLTWKSCSEYITRFL---DRGMGFLDIDIPKKKGLSKSVIAGVYD 473
QY 478 MDQCRDFGSGYQTCLEFRTEPCCKOLMCSHPDNPYFCGKTKKPPLDGTCEACAGKCKFC 537
Db 474 VHHOQOLGYPNATFCQEVENV--COTLMCSYKG---FCRSKLDAAADGTGCGKKKMA 528
QY 538 GHCI--WKSPEOTYQDGGMSWTFGSCSSCGGVASRSRSCNNSPAYGRRPLGPM 595
Db 529 GACITVKKKPEST---PGMGWMSHWSHSCSRCTGAGVQSAERLGNMPEPKFGKGYCTGER 585
QY 596 FEYQVONSEECPTGYEDFRADQCAKRNYSYVHONAKHSMWP--YPPDDAOCCELIQGSAD 654
Db 586 KRYRLCNVHPCSEAPTRFROMGCEFDI--VYKKNELVHMFITNP---AHPCELYCVRPID 641
QY 655 TGDVVFMMQVYHDTGRC--SYNDPYSVCARGCVPVGCDKEVYSKKADKCGVCGGDSHC 713
Db 642 GQFSKMLDAYIDGTCPCEEGNSRNVCINGICKMVGCDYEDISNATEDRCGVCLGDSGC 701
QY 714 RTVKGTLLKASKOAGALALVQIPAGARHIOLEALEKSHRSVYVANO-----VTGSTILN 767
Db 702 QVVR--KMKKQKSGSYVIGILPGARDIRMELEGAGNELIASEDPERYIYNGEFTI 760
QY 768 PKGKATSRRTTAMGLEWEDAVEDEAKESLKTGSPLEPAIAI--LALPTEGGPSSSLAYKY 826
Db 761 WNG-----NYKLGTAVYQYDRKGDLEKIMATPTNSVMIQLEFQYTING-----IKXY 810
QY 827 VTHEDLPLIGSNVNLLEEMTYEMALKSNAPSKACGGIOFTKYGCRRRRDHMYQRH 886
Db 811 TIQKGL---DNDV--EOM--YFWOYGHMTECSVTGCTGIRQTAHC--IKKRGWAKAT 861
QY 887 LCDHKRRKPIRRRCNOHPCSPQVWTEWAGACSRSCGKIGVOTRGIOCLPLISNGTHY 946
Db 862 FCDPEPTQNGRQKCKHEKAC--PRPMWAGEMWACATGPHGEKKRYVLCIQTWVSDQAL 920
QY 947 MPARACAGDREARPCLR--VPCPAORMLCAMSOCASVYCGEGIQORQVYCRITANSILGHC 1005
Db 921 PPTDCQHLLEKPKTLSCNRDILCPSDMVTGNMSECSVSCGGVIRISVTCANKHDE--PC 978
QY 1006 EGDR--PTVQVCSLPACG-----GNHONSTVADVWELCTEGCQVWPQSGPL 1051
Db 979 DVTKRPNRSLCGIQOCPSSRRVLPKPKKGTISGKNPPTLKPVPPTSRPRLTTPPTGPE 1038
QY 1052 HPINKISSMCAAEPCCT---GD-----RSVFCOMEYLDRYCSIPGYHR---LCVSCJIK 1098
Db 1039 SMSTSTPAISSPPTASKCDLIGKQWQSSQTPPELSSRYLSTSTSQPILTSLSLI 1098
QY 1099 KASGPV--PGDPPGTS---LPFSTGSPPLPGQDPAADAAPEPGKPT---GSEDHQH-- 1148
Db 1099 QPSEENWSSSDGTGPTSGGLVATTTSGSGLSSSRNPITPVPYFVNTLLKKGEMETHSGS 1158
QY 1149 GRATLPGALDTSPT-----GTQHPFAPE--TP--IPGASW--SISPTTIG 1188
Db 1159 GEEREQEDKEDSNPVIMTKIRVPONDAPVESHEMLAPPLPDLSSRESWMPFSTVWEG 1218
QY 1189 GLPMGWQTTPVPEDKGQGGEDLRIHG--TSLPAD--LP---GRPPECHPTGTFITLV 1241
Db 1219 LLP---SQRPT--TSETGP---RVEGMVTEKPAWTLPLIGDQHPSPSGKTANRNHLK 1269
QY 1242 LPRD 1245
Db 1270 LPNN 1273

```

```

ATSL_RAT
ID ATSL_RAT STANDARD: PRT: 967 AA.
AC 09WDQ1; Q9ERI1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 Precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the
thrombospondin type I motif (ADAMTS).";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luecht M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
MATRIX (BY SIMILARITY).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FORIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF149118; AAD34012.1; .
DR EMBL; AF304446; AAG29823.1; .
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; Tsp1.
DR InterPro: IPR000130; Zn_M12peptidse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; Tsp_1; 3.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS50215; ADAM_MERO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

```

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 54
 FT PROPEP 55 252
 FT CHAIN 253 967
 FT STATE 205 205
 FT METAL 401 401
 FT ACT_SITE 402 402
 FT METAL 405 405
 FT METAL 411 411
 FT DOMAIN 476 558
 FT DOMAIN 539 615
 FT DOMAIN 616 724
 FT DOMAIN 725 857
 FT DOMAIN 858 907
 FT DOMAIN 908 967
 FT DOMAIN 194 198
 FT CARBOHYD 547 547
 FT CARBOHYD 720 720
 FT CARBOHYD 764 764
 FT CARBOHYD 782 782
 FT CARBOHYD 945 945
 FT CONFLICT 21 21
 FT CONFLICT 26 31
 FT CONFLICT 49 49
 FT CONFLICT 72 72
 FT CONFLICT 79 79
 FT CONFLICT 249 249
 FT CONFLICT 262 265
 FT CONFLICT 607 607
 FT CONFLICT 936 936
 FT CONFLICT 962 962
 FT SEQUENCE 967 AA; 105705 MW; P93C864F6DCD4CF CnC64;

Query Match 16.1%; Score 1119.5; DB 1; Length 967;
 Best Local Similarity 29.2%; Pred. No. 3e-51;
 Matches 320; Conservative 151; Mismatches 393; Indels 231; Gaps 47;

QY 27 PELHL-SGKSLDYGVTVCS-----TDFRGRLSHVYSGRAASAGSMV----- 70
 DB 3 PEVPLGSGKLR-----PCSDMGDIQRAKFRSSOAHMLLLIALLTLLCVRGAGRP 56
 QY 71 ---DTPPLPRSHSLRVARSPRLRPGGTLMPGRVGRSHSLYENVYFGKELHLRLRNRRL 127
 DB 57 TDEDELVLPR-----SLERARGHDSITL-----LRDAFQQLHLKLQDPDSCF 99
 QY 128 VVPGSSVWQEDFRELFRQPLRQ-----ECYITGGVGMPCAAVAISNCGLAG 176
 DB 100 LAFGFTLQ-----TVGRSPGSEAOHLDPDGLAHCFYSGVNCDPSSAALSLCEGVRG 153
 QY 177 LIRTSIDTELEPR-----LER-----GGOEKEASGRTHVYVRR-----AVQOEPAE 218
 DB 154 AFYLOGEEFFIOPARAVATERLVAPKEESLAPRRPHILRRKRGSGGAKGVMBEETL 213
 QY 219 PDGDLHNEAFGLGDL-----PMLGLVGDOLG-DTERKRRHAKGYSIEVLVLVDSDV 272
 DB 214 PTNSNGRESQNPDPQWPLRNPPOG-AGKPTGPSIRKKRPVSSPRY-VETMLVADQDMA 271
 QY 273 RHGKEHVOYVLTLMNIVDELYADESGVINITALVRLINVGROSLIERNPBSRL 332
 DB 272 DRHG-SGLKHVLLTFLSVAAAREYKHPSTRNSISLVVKKILVI-YEOKGPEVTSNALTL 329
 QY 333 EGVCAWHAHQRODPS--HAHHHNVFLTRQDGRPS-----GYAPVTCCHPLRSCAL 384
 DB 330 RNFCSSW--OKOHNSPSSDRPDEHYTALLFTRODLGSHCTCOTLMDADVGYCDPSRCSV 387
 QY 385 NHEDGSAFVLAHETGVLGMEHD-----SQNGSCADETSIGSVMAPLVQAFHHPHWS 439
 DB 388 IEDDGQAFVTAHETGVLGMEHD-----SQNGSCADETSIGSVMAPLVQAFHHPHWS 444
 QY 440 RCKSLTSLRPLDS--YDCLLDPPRPAWQPELPLGINSWDEQCRFDESGGYOTCLAFR 497

DB 445 PCSAIVWTSFLDNGHCECLMDKRPQNI-KLPESDLPGTLYDANRQCFPFGESETHCP--D 501
 QY 498 TFEPCQOLWCSHEDNPF-CRTKKGPPDLDTGCAPGKWCFKGHCITWKPEDTQO--DGG 554
 DB 502 AASTCSTLMCTGSGGLLVCQTRKHPWADGTSCGSKKVCSCVKNTDMHFAFPVHGS 561
 QY 555 WSWMTFSGCSRSCGGGVSRSRSCNPPSPAYGPRCGLPMEYGVNSECSP-GTEYDF 613
 DB 562 WGFMPGWDGDSRSCGGGVSRYTHRECDNPVPKKGKCYCEKRRVRSCHIECPDNNKGTFF 621
 QY 614 RAQOCARNSYVYHQAOKS-----WVP-YEPDDAQKCELCIGSADTGDVYFMMOV 664
 DB 622 REQCEAHNPF-----SKASFGNEPTVEWTPKYAGVSPKDRCKLTCEAKGIGYFVLQPK 676
 QY 665 VHDGTRSYDPRYSVARGECYVGCDDKRVGSMKADDCGVGSDNSCRVYKTLGAKAS 724
 DB 677 VVDGTGPS-PDSTSVYVQGVCAAGCDRLTDSKKFKEDCKGCVGGNGSTCKKISGV--TS 733
 QY 725 KOAGATLVOIPAGARHIQTEALEKSPHRSVYKNQVTSF--ILNPKREATSRTTAMG 782
 DB 734 TRQYHDIYVTPAGATNIEVK-----HNPGRSNNGSFLAIRADGTYILNGFTLST 787
 QY 783 LEWE-----DAVEAKESLKTSGPLPEALIALALPPEGGPRSSIAKVIYIHDL 832
 DB 788 LEODLTYYKTVLRYSGSSALERISFSLKEPLTQVL----- 826
 QY 833 LPLGSNNVLLEMDTYEMALKSWAPCSKACGGGLOFTYKGRRRRHHMVQHLCDHK 892
 DB 827 --WVG-----HALRKIVYTF-----MKK 844
 QY 893 RPKRIRRRCHQPCSQPWWTEMGACSRSCGLGVQVTSIGQCLPLSGTHKVPAAKAC 952
 DB 845 KTEPF-----NAIPFSE-WVIEEMGDESKTGS-GWQRRVVEC--RDING--HPASEC 892
 QY 953 AGD-RPEARPRCLVPCPAOMRILGANSOCATGCEGIGQROVYVCFRTNANSLGHCEDGRP- 1010
 DB 893 AKEVPRPSTRCADLPP-RMVGDMSPGSKTGKGTKKRTKLKLSHDSGVLSTNSCPL 951
 QY 1011 ---DTVOVCSLPAC 1021
 DB 952 KRPKHVTDICILITQC 966

RESULT 8
 ID ATSL_MOUSE STANDARD; PRT; 968 AA.
 AC P97857; 054768;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
 GN ADAMTS1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN NCBI
 RP STRAIN=129/SVJ;
 RC MEDLINE=98110583; PubMed=9441751;
 RA Kuno K., Iizasa H., Ohno S., Matsushima K.;
 RT "The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
 RL Genomics 46:466-471(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97150761; PubMed=8995297;
 RA Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F., Matsushima K.;
 RT "Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene.";

RL J. Biol. Chem. 272:556-562(1997).

RN [3]

RP CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.

RX MEDLINE=99303657; PubMed=10373500.

RA Kuno K., Terashima Y., Matsushima K.;

RT "ADAMTS-1 is an active metalloproteinase associated with the extracellular matrix."

RL J. Biol. Chem. 274:18821-18826(1999).

RN [4]

RP FUNCTION.

RX MEDLINE=20389568; PubMed=10930576;

RA Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.;

RT "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan."

RL FEBS Lett. 478:241-245(2000).

RN [5]

RP FUNCTION, AND INDUCTION.

RX MEDLINE=20243757; PubMed=10781075;

RA Rodger R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W., Richards J.S.;

RT "Progestosterone-regulated genes in the ovulation process: ADAMTS-1 and cathepsin L proteases."

RL Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).

CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-1-LEU-1692 SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.

CC -1- CORPCTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX.

CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA CELLS OF PREOVULATORY FOLLICLES.

CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -1- PFM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 7.

CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AB001735; BAA24501.1; ALT_INT.

DR EMBL: D67076; BAA11088.1; ALT_FRAME.

DR MEROPS: M12.222; .

DR MGD: MGI:109249; Adamts1.

DR InterPro: IPR001762; Disintegrin.

DR InterPro: IPR002870; Pep_M12B_propep.

DR InterPro: IPR001590; Reprolysin.

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR000130; Zn_Mtpeptidase.

DR Pfam: PF01562; Pep_M12B_propep; 1.

DR Pfam: PF00090; TSP1; 3.

DR SMART: SMO0209; TSP1; 3.

DR PROSITE: PS0215; ADAM_MEROPS; 1.

DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.

DR PROSITE: PS0092; TSP1; 3.

DR PROSITE: PS00142; ZINC_PROTEASE; 1.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat: Extracellular matrix; Heparin-binding.

FT SIGNAL 1 48

FT PROPEP 49 253

FT CHAIN 254 968

FT SITE 206 206

FT METAL 402 402

FT ACT_SITE 403 403

FT METAL 406 406

FT METAL 412 412

FT DOMAIN 477 559

FT DOMAIN 560 617

FT DOMAIN 618 725

FT DOMAIN 726 850

FT DOMAIN 851 909

FT DOMAIN 910 968

FT DOMAIN 195 199

FT CARBOHYD 548 548

FT CARBOHYD 721 721

FT CARBOHYD 765 765

FT CARBOHYD 783 783

FT CARBOHYD 946 946

FT MOTIFEN 403 403

FT MOTIFEN 335 335

FT CONFLICT 425 425

SO SQUENCE 968 AA; 105841 MW; 42EBDA35499FB6C1 CRC64;

Query Match 15.9%; Score 1108.5; DB 1; Length 968;

Best Local Similarity 29.2%; Pred. No. 1,1e-50;

Matches 300; Conservative 141; Mismatches 364; Indels 223; Gaps 41;

Db 87 RSLHGGGLMPGRGRHSLYNVTVFGEHLRLRPNRRLVPGSSVEMOEDFRELEFRQ 146

70 RAGHDSTT-----TRLRLDAFGQDLTKIQDPSGFALPFTLQ-----TVGRS 113

147 PLRQ-----ECVYTGCVTPGPAVAVNSMCDLGLIRIDSTDIETP----- 189

114 PGEADHDPTDGLMCHFTSGTVNGDPSAALSLCEVRAFAFYQGEFFIOPAPGVAT 173

190 --LERGOEKEASGRT--HVYVRE-----AVQGMARPDGLINEA----- 227

174 ERLAPVAPREESSARPROFILRRRRSGGAKCGMDETLPTDSRESQNTNOMPRV 233

228 -----FGLGLPNLGLVGQLDTEKRRHNAKPGSYSTEVLLVYDVSVPFNGK 277

234 DTPPDAGKPSGSGSI-----RKKEFVSSPRY-VETMLVADQSNADPHG- 276

278 EHYONYVLTINIVDEIYHDESLGVHINIALVRLIMVGYRQSLILIRGNPSRLGEVCR 337

277 SGLKHYLLTFVAAARFKHPSIRNSISLVYVKIIVI-YEQKGREVTSNAALTLRNFEN 335

338 MAHSQORODPS--HAENHDHVVFLTRDQFGPS-----GVAPTYGMCHPLRSCALNHEDG 389

336 W--ORQHNSPSDRDEPHDFTAILFTRDLCGSHNCDTLGMADVCTVDDPSKSVIEDDG 393

390 FSSAFVIAHETGHVGMHDGNGCAD---ETSLGSMAPLVQAAHRRHNRWSCKTEL 446

394 LOAATTTTAEHGLHVFNMHD--DAKHCASLNGVTDGSHLMSMSLSDHSOPWSPSCSAUYW 452

447 SKYLP--YDCLDDPPDPAMPORPELPGIINYSMDEOCRFDFGSGYGTCLAFTRFPCQ 504

453 TSFLNGHGECILMDKRPQPI-KLPISDLPGTLVDANRCQQTTFGESEKHC--DAASTCTT 509

505 IMCSHPDMPYF-CKTKKRPPLDTEGCAPKWCXGKCGHGTMSPEQYGO--DGGWSWTFK 561

510 LMCSTGSGLLVCTKHPWADGTSCEGKWCVSGKVNKTDKHFATPVHSGWGPWGW 569

562 GSCSRGCGGVRSSRCNNPSPAYGRRPGLGPMFEYOVONSSECP-GTYEDPRAQCK 620

570 GDSGTGCGGQVYTMRECDNAPVKNNGKYTEGKRVKRRSNIEDCDNNNGKTFREQCEA 629

621 RNSYTVHOMAKHS-----WVP-YEPDDAQKCELCISADTGDGVVPMNQVHDTGRC 671

Db 630 HNEF-----SKASFGNEFTVEMTPRYAGVSPKDRCKLTCEAKGIGYFVLIQPKVVDGTPC 684
 QY 672 SYRDPYVCAGBECVPCVQKREVSQMKADPCQYGGGDNHCHPTVKGTLGKASQKAGALK 731
 Db 665 S-PDSTSVCGVQGCYKACCDRIIDSKKFFEDCQYGGGSGCKMAGSYV--TSRPGYHD 741
 QY 732 LVQIPAGARHIOIEALEKSPHRSVYKNOVTSF--ILNPKCKEATSRFTFMGLMEF--- 786
 Db 742 IYTIIPAGATNIEVK-----HRNQRGSRNNSFLAIRAADTYILLNGFTLSTEDQDITY 795
 QY 787 -----DAVEDAKESLTSQGPLPEAIVALLPTEGGPRSSLAKEYIHEDLLPLIGSN 839
 Db 796 KGTVLRYSQSSAALERINSFSLKPELITQV-----MMG-- 830
 QY 840 NVLLEEMDYEWALKSMAPCSKACGGGIGFTKYGCRRRRDHMHVORHLCDHKRPPKPIRR 899
 Db 831 -----HALRPKIKFTYF-----KKKTESF-- 850
 QY 900 RQNHPCSPVAVTEEMGACSCGCKLGVQRTGICLLPLSNGTHKVPAPAKACAGD-RPE 958
 Db 851 --NAIPTESE-WVIEWEGSCSKTGS-GWQRVYVC--RDING---HPASCAKEVKA 900
 QY 959 ARPPCLRVCPAPWRLGAWSCSATCGEGIGQRYVYCTNANSLGHCEDRP-----DTV 1013
 Db 901 STRPCADLPCP-HMVGDMSPSCSKTGKGYKRTIKVYSHOGVLSNEDPLKKPKHYI 959
 QY 1014 QVCSLPAC 1021
 Db 960 DECTLTQC 967

RESULT 9
 ATSL_HUMAN STANDARD; PRT: 967 AA.
 ID ATSL_HUMAN Q9UHB8; Q9UHB3; Q9P2K0; Q9NSU8;
 AC Q9UHB8; Q9P2K0; Q9UHB3; Q9NSU8; Created
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
 GN ADAMTS1 OR METH1 OR KIAA1346.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
 RA "Cloning, characterization and mapping on human chromosome 21 of the
 RT orthologue of murine Adamts-1.";
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A., AND FUNCTION.
 RA TISSUE=Heart;
 RA MEDLINE=99367466; PubMed=10438512;
 RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
 RA Lombardo M., Iruela-Arispe M.L.;
 RA "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
 RT family of proteins with angio-inhibitory activity.";
 RT J. Biol. Chem. 274:23349-23357(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Endothelial cells;
 RA MEDLINE=20247184; PubMed=10785405;
 RA Glennie J., Schmitt A.O., Pilarsky C., Hinzmann B., Weiss B.,
 RA Rosenthal A., Thierach K.H.;
 RA "Differential gene expression by endothelial cells in distinct
 RT angiogenic states.";
 RT Eur. J. Biochem. 267:2820-2830(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Brain;
 RA MEDLINE=20181126; PubMed=10718198;

RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RL for large proteins in vitro.";
 RN DNA Res. 7:65-73(2000).
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totski Y., Choi D.-K., Soeda E.,
 RA Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordstam G., Hornischer K., Brandt P.,
 RA Schaefer M., Schoen O., Desario A., Reischelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Resselmann L., Dagand E.,
 RA Wehrmeyer S., Bozzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RN Nature 405:311-319(2000).
 RP SEQUENCE OF 418-967 FROM N.A.
 RA TISSUE=Melanoma;
 RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
 CC ACTIVITY. ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH
 CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
 CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
 CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-1-LEU-1939
 CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 CC MATRIX (BY SIMILARITY).
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: AF170084; AAF15317.1; -
 DR EMBL: AF060152; AAD48080.1; ALT_INIT.
 DR EMBL: AF207664; AAF23772.1; -
 DR EMBL: AB037767; BAA92584.1; ALT_INIT.
 DR EMBL: AP001697; BAA95502.1; -
 DR EMBL: AL162080; CAB82413.1; -
 DR MIM: 605174; -
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR002870; Pep_M12B_propep.
 DR InterPro: IPR001590; Reptolysin.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01562; Pep_M12B_propep; 1.
 DR Pfam: PF01421; Reptolysin; 1.
 DR Pfam: PF00090; tsp.1; 3.
 DR SMART: SM00209; TSP1; 3.
 DR PROSITE: PS50215; ADAM_MEPRO; 1.
 DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR PROSITE: PS50092; TSP1; 2.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.

KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KM Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 49 POTENTIAL.
 FT PROPEP 253 252 BY SIMILARITY.
 FT CHAIN 253 967 ADAMTS-1.
 FT SITE 198 198 CYSTEINE SWITCH (POTENTIAL).
 FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 402 402 BY SIMILARITY.
 FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 476 559 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 560 616 DISINTEGRIN-LIKE.
 FT DOMAIN 617 724 TSP TYPE-1 1.
 FT DOMAIN 725 849 TSP TYPE-1 2.
 FT DOMAIN 850 908 TSP TYPE-1 3.
 FT DOMAIN 909 967 TSP TYPE-1 3.
 FT DOMAIN 843 846 POLY-LYS.
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 227 227 P -> A (IN REF. 4 AND 5).
 FT CONFLICT 468 468 O -> H (IN REF. 1).
 FT CONFLICT 561 561 S -> N (IN REF. 1).
 SQ SEQUENCE 967 AA: 105383 MM: C189389324741BD1 CRC64;

Query Match 15.8%; Score 1098.5; DB 1; Length 967;
 Best Local Similarity 29.2%; Pred. No. 3.8e-50;
 Matches 300; Conservative 145; Mismatches 361; Indels 223; Gaps 46;

QY 87 RSPPLPGTLMPCRYGRSLFNTVYEGKELRLRPNRRLVPGSSVEQEDFRELFRQ 146
 DB 67 RAPGH--GTT--RLRLHA-----FDQDLLELRDSSFLAGFTLQ--NVGKKSSST 113
 QY 147 PLRQ-----ECYVTGVTMPCGAAYAINSCGLIITDSDFTPE----- 189
 DB 114 PLPELDLHACRYSGTVNDPSSAALSLCEBVRGAFYLLGEAFYTOPLPASERLATAP 173
 QY 190 -----LERGOQ-----EKEASGRTHVYRREAVQOE-----NAEPD 220
 DB 174 GEKPPAPLQFLHRLRRNRGQDVGTCGVVDEPRPTKAEFEDEDEGEDEGPQMSPOD 233
 QY 221 GDLHNEAFGLDLPNLGLVGDQIG--DTEKRRRAKRGSTIEVLYVDDSVYRFGKEH 279
 DB 224 -----PALGO--VGQPTGTGSIKKRKFVSSHRY--VETMLVAVDQSMAPFEG--SG 277
 QY 280 VQNTVLTLMNIVDEIYHDESLGVHINIALVLLVAVGYSRLSLIERGNPERSLEYQCMA 339
 DB 278 LKHYLLTLEFVAARLYKHPISIRNSVSLVAVKLLVIHDEQKGPBY--ISNALTLIRNFCNM-- 335
 QY 340 HSGQRDPS--HAHHDHVFLTRQDGPB-----GYAPVTGMCHPLRSCALNHEDGFS 391
 DB 336 -OKOHNPSSDRADEHYTALILFTRODLCGSQCTDLGMDVGVTCDPSPRSQVIEDDQ 394
 QY 392 SAFVIAHETGTVLMEHDGCGNCADETSLGS--VWAPLYQAARFRHMSRCKIELSR 448
 DB 395 AAFVTAHELGVFMRPD--DAKOCASLNGVNOSSHMAKSLMLDSQSPSPSAVMTYS 453
 QY 449 YLPS--YDCLLDPEPDPAPQPELPQINSMDQCRPFGSGYQTCIAFRTFEPCKOLM 506
 DB 454 FLNDGHECLMDKFOPEI--QLPGDLFTSYDANRQCOFFTEGDSKICP--DAASTCSTIM 510
 QY 507 CSHNDNYF--CKTRKGPRLTEGCACAGCKCFKHCWKSPEDQYGO--DGMSSWTKEGS 563
 DB 511 CTGSGSVLYCQTKHFMADGTSKCGEKKWCTINKCNKTKRKHFTDPPFGSGMMGMPMD 570
 QY 564 CSRSGGVSRSRSCNPSPAYGPRCLGPMFEYOVNCSSECP--GYEDFPAOOCARKN 622
 DB 571 CSRTCGGVOYTMRECDNPVKKNGKYCEBKRYRNSCULECPDNNKGTTFEEQCEAHN 630
 QY 623 SYVYHNAKHS-----VYF--YEPDDDAQCKELICQSLADGVDVFNQVVDHGTGRCY 673
 DB 631 EF-----SKAFSGPAVEWIPRYAGVSPRDKCLICQAKGIGYFVLIQPKVVDGTPCS- 684

QY 674 RDPYVCARGECYVPCDKEVSGMKADKCGVCGGDNHCRFTVGTGLKASKOAGALKIV 733
 DB 685 PDSVYCVGCGCKVCKDRILIDSKKKFKDCKGCGGNGSTCKKISSV--TSAKGYHDIT 742
 QY 734 QIPAGRHQIEALEKSPHRS-----VKNQYTGSIPLPKKEATSRFTAGLEWE--- 786
 DB 743 TLPGTATNIEVORNRQGRSRRNNGSFLAIKADGYIIL-----GDYTLSTLEQDIMY 794
 QY 787 -----DAVEDAKSLKSTGPLEPAIALIPLPTEGEP--RSSIATYVIVHEDLLPLIGS 838
 DB 795 KGVLYRYSSSSALERINSFSPLEKPTIYVL--TVGALPKIKITYYV----- 842
 QY 839 NNVLLEMDYEMALKSWAPCSKACGGIOFTKYGCRRRRDHMYORHLCDHKRPKPIR 858
 DB 843 -----KKK-----K 846
 QY 899 RRCNOHPCQPYWYTEEMGACSRSGKIGVOTRGIQCLLPISNTHKVPKAKACAD-RP 957
 DB 847 EEFNALP--TFSAWVTEEWGECSSKSC--ELGWQRRLVEC--RDING-----OPASECAKEVKP 898
 QY 958 EARRPCLRPYCPAOWRLGAMSCSATCGEIGIOHQCRT--NANSLGHCSD--RP-DT 1012
 DB 899 ASTRCADHPCP--QWQJGEMSSCSKTKGKTKKSLACLSDHGCVLSHESCDPLKPKHP 957
 QY 1013 VOYCSLPAC 1021
 DB 958 IDECTMACC 966

RESULT 10
 AT55_HUMAN
 ID AT55_HUMAN STANDARD; PRT; 930 AA.
 AC Q9UNA0; Q9URP2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-5 precursor (EC 3.4.24. .) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
 DE (ADMP-2) (ADAM-TS 11).
 GN ADAMTS5 OR ADMP2 OR ADAMTS11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=99367476; PubMed=10438522;
 RA Abdazade I., Liu R.-O., Yang F., Rosenfeld S.A., Ross O.H., Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M., Wynn R., Duke J.L., George H.J., Hillman M.C., Jr., Murphy K., Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H., Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F., Arner E.C., Burn T.C.;
 RT "Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family";
 RL J. Biol. Chem. 274:23443-23450(1999).
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohtsuki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W., Rosenthal A., Kuschi T., Shibuya K., Kawasaki K., Asakawa S., Shimizu A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E., Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P., Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E., Wehner S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt K., Yaspo M.-L.;

RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN (3)
 RC SEQUENCE OF 413-930 FROM N.A.
 RP TISSUE-Fetal brain;
 RA MEDLINE-9395124; Pubmed-10464288;
 RX Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
 zinc metalloproteases.";
 RL J. Biol. Chem. 274:25555-25563(1999).
 CC -1- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 DESTRUCTION OF AGGRECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
 PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
 CC -1- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 392-GLU-1-ALA-393
 SITE.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 MATRIX (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
 BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
 CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
 CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
 ARTHRITIC PATIENT.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PFM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announce/>
 or send an email to license@isb.sib.ch).
 CC -----
 CC EMBL: AF142099; AAD49577.1; -
 DR EMBL: AP001698; BAA95504.1; -
 DR EMBL: AP001697; BAA95503.1; -
 DR EMBL: AF141293; AAF02493.1; -
 DR HSSP: Q9PW35; 1BDJ.
 DR MIM: 605007; -
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR001590; Repolysin.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR000130; Zn_MTPeptide.
 CC Pfam: PF01421; Repolysin; 1.
 CC Pfam: PF00090; TSP_1; 2.
 CC SMART: SM00209; TSP1; 2.
 CC PROSITE: PS50215; ADAM_MPRO; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC PROSITE: PS00092; TSP1; 1.
 DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 KM Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; zymogen;
 KW Repeat; Extracellular matrix.
 FT SIGNAL 1 16 POTENTIAL.
 FT PROPEP 17 261 POTENTIAL.
 FT CHAIN 262 930 ADAMTS-5.
 FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
 FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 485 566 DISINTEGRIN-LIKE.
 FT DOMAIN 567 623 TSP TYPE-1 1.
 FT DOMAIN 624 731 CYS-RICH.
 FT DOMAIN 732 874 SPACER.
 FT DOMAIN 875 930 TSP TYPE-1 2.
 FT DOMAIN 37 41 POLY-ALA.

FT DOMAIN 257 261 POLY-ARG.
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 138 138 A -> G (IN REF. 2).
 FT CONFLICT 614 614 R -> H (IN REF. 3).
 FT CONFLICT 692 692 P -> L (IN REF. 2).
 SQ SEQUENCE 930 AA; 101715 MW; B64281502F28193B CRC64;

Query Match 15.4%; Score 1068; DB 1; Length 930;
 Best Local Similarity 30.0%; Pred. No. 1.4e-48;
 Matches 309; Conservative 119; Mismatches 362; Indels 240; Gaps 43;

59 GPAASAGSVVDPPTL-----PRSHSLVY--ARSLPRL----- 92
 21 GPAATPADK-AGOPTAAMAAPRRROGEVEVERAEPGPHPLAORRSKGLVONIDQ 79
 93 ---GGTLPNGRYGRSLVENYVFGKELRLRPRL-----VVP---GSSVEQEDFRE 142
 80 IYSG---GKVG-----YLYAGRRFLDLERGSVINGFVPAAGGTAPMRH----- 126
 143 LFRPLRQECVYTVGVGMGAANAISNCDGLIRFDSDFTEPLERQOQKEASGR 202
 127 -----RSHCFYRGTVDSFSLAVFDLCGGIDGFFAAVKAHAYTLPLRPMWAEKGR 180
 203 TH-----VYKREAVQDEAEFDG-----DLHNEAFGLGD.PNLLGLVGDQ 245
 181 VYGDGSAFRLHYTRGFSEFELPPASCEPASTPEHEHAPSHN-PSGRALASDL 239
 246 DRE-----RRRAKPGSYSIEVLAVDVSVFNGKEHYONTVLTAMIV 291
 240 DOSALSPAGSGPQTMWRKRRRSISRAQVELLIVADSMARLYR-GLOHYLTASIA 298
 292 DELYHDESGVHINIALVRLWVGRQSLIERGNPSLSLQVCRMAHSQOROPSAE 351
 299 NRLYSHASIEHNLRLAVVAVVYVLDGDK-SLEVSNAATLTKFKWQHONQDDBHE 357
 352 HHNVHVELRQD-----GPGSVAPVTGMCBPLRSLNHDGSSAFVIAHETGHLV 405
 358 HYDAILFTREDLGGHSCDTIGMADVGTICSPERSCAVIEDDGLHAANTVAHEGLHLG 417
 406 MEHDGONGCADETSLSG-----VNAPIVQAAPHRFMSRCSKLSRYLPS--YDCILD 458
 418 LSHD-DSKFC--EETFGSTEDKRLMSITLSDASKPMKCTSAVTEELDGHGNCILD 474
 459 DPDPAPMPQPELPLGINSDEQCRDRPGSGYOTCLAFTEFPCQQLWCS-HPMDPYECK 517
 475 LPRKQIL-GEELPQGTVDATQCNLTFGREYVCPG--MDVCARLMCAVVRQGMVCL 530
 518 TRKGPPLDTGECAPGKWCFCGHCIMKSPDQY--GQDGGMSWTKRGSCSGGGVRSR 575
 531 TKKLPAVEGTPCGKGRKCLQKGVCDTKKKYTSSSHGMWGSWGSGSCGGGVQPA 590
 576 SRSCNNSPAYGRPLGPMFEYQVNSSECPGTIEDFAQOCARKNSTYVQNMKH--- 632
 591 YRHCHNPNAPRNNRGYCTGKRAIYRSCSLMPCPNKGSFRHEDCEARNGY--OSDAKYKT 648
 633 --SWVP-YEPDDAOKCELCOSADTGDVYFMQVYVHDFGRCSYRPY--SYCARGECEVP 687
 649 FVEWPEYAGVLDADYCKLCKRAKGTGYVYVSPKTYDTGECC--PYNSNCVARKCYVR 705
 688 VGCDEKESMKADKCGVCGGDSHCRVYKGLGKASKAGALKVLQIPAGARHIQIE-- 745
 706 TGCDGIIIGSLQYDKCGVCGGDSCTKIYGTENKSK--GYDVVRIDEGATHIKVRF 763
 746 -----ALEKSPHRSYVKNQYTGSEFLNPKGKATSTF-----TAMGLMEDA 788
 764 KAKDQTRFAYVLAK-----KN-----GEYLLNGRYMISTETIINDINGTVNNGWSH 813
 789 VED-----AKESLKTSGPLPEAIALALPTEGGRSSLAAYKVIHEDLLPLIGSN 839

```

Db 814 RDDFLHGMGYSATKEIL-----IVQILANDPTK---PLDVRYSEFVPKKSTPKVNS- 861
QY 840 NVLLEEMDTYEWALKSWAFCSKACGGGIOFTKYGCRRRRDHMHVQRLCDHKRKPPIRR 899
Db 862 -----VTSHGSKNYGSH----- 873
QY 900 RCNQHPCSQPVWVTEBEWGCASRSCGKLGVOYOTRGIOCLPLSNGTHKYMPAKAC-AGDRPE 958
Db 874 -----TSQPOWVTGFWLACSRIC-DTGWHTRTYVC-----QDGNRKL--AKGCPLOSQPS 920
QY 959 ARRPCLRVPC 968
Db 921 AFGQCLLKKC 930

```

Job completed: August 28, 2002, 20:14:54
 Elapsed time: 261 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 19:52:08 ; Search time 66.13 Seconds
(Without alignments) 2102.894 Million cell updates/sec

Title: US-09-938-330-20

Perfect score: 6954
Sequence: 1 MAPLEALLSYLLPLHCALCT.....PTGTFLLCVLPDSDQLNGHT 1252

ng table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A Geneseq 032802.*
2: /SID5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SID5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SID5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5160.5	74.2	1048	22	AAB85695	Larynx carcinoma e
2	3624	52.1	1205	22	AAB73550	Human ADAM-type me
3	3614	52.0	1201	21	AAB21254	Human metalloprotei
4	3327	47.8	1203	22	AAB50004	Bovine metalloprotei
5	3299.5	47.4	1211	19	AAM47028	Human N-proteinase
6	3279.5	47.2	1328	22	AAB22373	Novel human diagno
7	3270.5	47.0	1206	19	AAM47030	Bovine N-proteinase
8	2962.5	42.6	969	21	AAY53900	Amino acid sequenc
9	1647	23.7	566	19	AAM47029	Human N-proteinase
10	1438	20.7	1686	22	AAB74944	Human ADAM type me
11	1436	20.6	1686	22	AAE00913	Human 27875 ADAM-T

12	1435	20.6	1686	22 AAE00934	Human 27875 ADAM-T
13	1426	20.5	1690	22 AAB6949	Human metalloprote
14	1415	20.3	1784	21 AAB41379	Human ORFX ORF1143
15	1328	19.1	1103	22 AAB74945	Human ADAM type me
16	1240.5	17.8	1934	22 AAB72301	Human ADAMTS-9 alt
17	1221	17.6	1072	22 AAB72300	Human ADAMTS-10 al
18	1220.5	17.6	1081	22 AAB72288	Human ADAMTS-10 al
19	1216	17.5	1073	21 AAB21264	Human metalloprote
20	1164.5	16.7	959	22 AAB86947	Human metalloprote
21	1159.5	16.7	908	22 AAE03572	Human metalloprote
22	1147.5	16.5	1882	22 AAB72286	Human ADAMTS-9 am
23	1121.5	16.1	958	21 AAB21255	Human metalloprote
24	1118	16.1	950	21 AAB62399	Human ADAMTS-9 am
25	1117.5	16.1	874	22 AAB72287	Human metalloprote
26	1110.5	16.0	947	22 AAB86950	Human metalloprote
27	1102.5	15.9	968	22 AAB50011	Human metalloprote
28	1098.5	15.8	950	20 AAY49501	Human ADAMTS-9 am
29	1098.5	15.8	950	22 AAB73549	Human ADAMTS-9 am
30	1098.5	15.8	950	22 AAB50002	Human ADAMTS-9 am
31	1098.5	15.8	967	19 AAB80285	Human ADAMTS-9 am
32	1098.5	15.8	967	20 AAY04142	Human ADAMTS-9 am
33	1088.5	15.7	967	20 AAY78189	Human ADAMTS-9 am
34	1079.5	15.5	950	21 AAY53899	Human ADAMTS-9 am
35	1075	15.5	2150	21 AAY53898	Human ADAMTS-9 am
36	1072.5	15.4	2165	22 AAB90617	Human ADAMTS-9 am
37	1068	15.4	930	20 AAB75426	Human ADAMTS-9 am
38	1053.5	15.1	997	22 AAB72283	Human ADAMTS-9 am
39	1052.5	15.1	891	22 AAB41226	Human ADAMTS-9 am
40	1052.5	15.1	929	21 AAB72282	Human ADAMTS-9 am
41	1040	15.0	859	22 AAB72282	Human ADAMTS-9 am
42	1038.5	14.9	1054	22 AAB60410	Human ADAMTS-9 am
43	1037.5	14.9	757	22 AAE03583	Human ADAMTS-9 am
44	1032	14.8	896	21 AAB21265	Human ADAMTS-9 am
45	1025	14.7	930	22 AAB72280	Human ADAMTS-9 am

ALIGNMENTS

RESULT 1	
ID AAB85695	standard; Protein; 1048 AA.
AC AAB85695;	
XX	
DT 29-OCT-2001	(first entry)
XX	
DE Larynx carcinoma associated protein-1 (LarCAP-1) polypeptide.	
XX	
KW Larynx carcinoma associated protein-1; LarCAP-1; carcinoma; cytostatic;	
KW antineoplastic; antiarthritic; osteopathic; immunosuppressive; human;	
KW immunostimulant; cerebroprotective; vasotrophic; antiinflammatory;	
KW dermatological; cardiac; vaccine.	
XX	
OS Homo sapiens.	
XX	
PN WO200159133-A1	
XX	
PD 16-AUG-2001.	
XX	
PF 12-FEB-2001; 2001WO-EP01525.	
XX	
PR 14-FEB-2000; 2000EP-0102955.	
XX	
PA (MERRE) MERCK PATENT GMBH.	
XX	
PI Duecker K, Heintsch B, Hoheisel J, Frohme M;	
XX	
DR WPI: 2001-483569/52.	
XX	
DR N-PSDB: AAM47048.	
XX	
PT Larynx carcinoma associated protein-1 polypeptide for the treatment of	
PT carcinoma, metastasis, arthritis, osteoporosis, immune disorders,	

stroke, ischemia, autoimmune disease, angiogenesis, skin disorders and organ malfunctions -

Claim 1: Page 33-39; 47pp; English.

This represents a human larynx carcinoma associated protein-1 (larcap-1) polypeptide. The larcap-1 polypeptide can be expressed by standard recombinant methodology. The larcap-1 polypeptide and polynucleotide are useful in diagnostic assays and for the treatment of carcinomas, metastasis, arthritis, osteoporosis, immune disorders, stroke, ischemia, autoimmune disease, angiogenesis, skin disorders and organ malfunctions especially heart hypertrophy.

Sequence 1048 AA:

Query Match 74.2%; Score 5160.5; DB 22; Length 1048;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 925; Conservative 0; Mismatches 6; Indels 39; Gaps 2;

291 VDEYHDESGVHNIALVRLVWGYRQSLTERGNPSKSLBOVGRMAHSDQRODPSSHA 350
80 vdeyhdeslgvhnialvrlwgyrqslltergnpsrslqevcrwahsgqrdpsna 139
351 EHHDDVFLTRDGFSGYAPVTGMCHPLRSALNHEGSSAFVIAHETGHVLMGHDG 410
140 ehhdvfltrdgtfsgyapvtgmchplrsalnhegssafviahetghvlgmehdg 199
411 OGNCACBETSGVMAPLNOAAPHFHMRSKLELSRYLPSPDCLDDDFDPAWOPPE 470
200 qgncacbetstgvmaplnoaaphfhmrsklesrlylpspdclddpfdpawoppe 259
471 LPGINSKDEOCRPDFSGYQTLAFTFBPCQQLWMSHPDNPFCKTKKGPPLDGETCA 530
260 lpginskdeocrpdfsgyqtlaftrfbpcqqlwmschpdpnfcktkkppldgetca 319
531 PGKCFKHCITWKSPEQTYGDDGMSWTKRGSCRSRSGGVRSRSCNNPSPAYGGR 590
320 pgkcfkchcwtkspeotygdgmswtkrgscrsrsggvrsrcnnpsayggr 590
320 pgkcfkchcwtkspeotygdgmswtkrgscrsrsggvrsrcnnpsayggr 590
591 CLGPFEEYQVQCNSECEGTGYEDFRAOOCARNSYTVHQNKHSMWPEPDDDAOKCELIC 650
380 clgpfefeyqvqcnsecegtgyedfraoocarnsytyvhnkshsmwpepdddaokcelic 439
651 QSADTGVFVNNVYHDCSTYRDPYSVCARGCVPVCGDKKESGMKADKCGVCGGDN 710
440 qsadtgvfvnnvyhdcstyrdpysvcargcvcvpgcdkevsgmkadckgcvggdn 499
711 SHCRTVKTGLKASKOAGALKIVQIPAGARHIQIEALEKSPRSVKNNOVTSFILNPKG 770
500 shcrtvktglkaskoagalkivqipagarihiealeksprrsvknnovtspfllnpgk 559
771 KEATSRFTFANGLEWEDAVEDEAKESLKTSGPLPEALIALPPEEGRRSLAKKYIHE 830
560 keatsrftfanglewedavedekeslktsgplpealialppeegrrslakkyihe 619
831 DLPLIGSNVYLEMDYIEMALKSMAPCSKACGGGIFTYTCGRRRRDHMMVRRHLCDDH 890
620 dlpligsnvylemdyiemalkswapcskacgggiftlytcgrrrrdhmmvrrhlcddh 679
891 KRRPPIRRRCNOHPCSOPVWTEEMGACSRSCGLVQVNGIQCLPLSLNGTHKVPKAK 950
680 krrppirrrrcnohpcsopvwtteemgacsrscglvqvnqiqclplslngthkvpak 739
951 ACAGRPPEARRRCLAVPCPAOMRLGANSO----- 979
740 acagrppearrrrclavpcpawrlgawsgkylstscmpdlvlmrepsingteliallv 799
980 -----CSATCGEIGIOOROVVCRRTNANSLGHCEDRPTVOYCSLPACGNGHONSTVRADV 1034
800 gptvicsatcgsglqirgvcrtnanslghecdpdrptvcvcslpacgngnqstvradv 859
1035 WELGTPPEGQWVDPQSGPLHPIKISSMCAABPCTGDRSVFCQMEVLDRCISGTHRLCCV 1094

|||||
860 wclgtpeqwwpvgsephlpinklss-----tepcvgdsvicqmevldaycsipgyhnlccv 916

1095 SCIKKASGPNPDPGPPTSLPPEFSTPGSPLPGDPADAAEPGKPTGSEDOHGRATOL 1154

917 scikkaasgpnnpdpgpptslppefstpgsplpgdpadaaepgkptgsebhgratql 976

1155 PGALDTSSPGTOHPFAPETPIPGASWSISPTTPEGGLPMWMTQPTTPYPERKGGPGEELRH 1214

977 pgaldtsspgtqhpfapetpfigaswsispttpgylpwmtqpttpyepkqggedlrlh 1036

1215 PGTSIPADLP 1224

1037 pqtstipaasp 1046

RESULT 2

AAB73550
ID AAB73550 standard; Protein: 1205 AA.

AAB73550;
07-AUG-2001 (first entry)

Human ADAM-type metalloprotease MDR5, SEQ ID NO:10.

Human; MDR5; ADAM-type metalloprotease; drug screening;

A Disintegrin And Metalloprotease; cancer; arthritis.

Human sapiens.

JP2001017183-A.

23-JAN-2001.

09-JUL-1999; 99JP-0196584.

09-JUL-1999; 99JP-0196584.

(YAMA) YAMANOCHI PHARM CO LTD.

WPI: 2001-275950/29.

N-PSDB: AAB70225.

PT

A new metal protease and its preparation for use as an anti-cancer and anti-arthritis therapeutic -

Claim 1: Page 16-19; 22pp; Japanese.

The invention relates to the novel human ADAM (A Disintegrin And Metalloprotease)-type metalloproteases MDR5 (AAB73549) and MDR55 (AAB73550). The metalloproteases can be used for the treatment of cancers and arthritis. The invention also relates to the genes encoding MDR5 and MDR55, vectors and host cells containing the MDR5 or MDR55 genes, the recombinant production of MDR54 and MDR55, and antibody specific for MDR54 or MDR55, and methods of screening for compounds which modulate the activity of MDR54 and/or MDR55. The present sequence represents human MDR55.

Sequence 1205 AA;

Query Match 52.1%; Score 3624; DB 22; Length 1205;
Best Local Similarity 55.6%; Pred. No. 2.8e-249;
Matches 666; Conservative 145; Mismatches 235; Indels 152; Gaps 17;

37 DYGVTPCSTDFRGRFLSHVSGPAAASGSMVVDPPPLPRHSHLR-----VARSPH 91

41 eyelvtpvcsnldgrflshls-----ashktrsdvsnp-- 77

92 PGGLTWGVRGSRSLVFNVTVEGKELHLRLPNRRLVVGSSVEMOE----- 138

78 -----eqllfnltafgkdfnlrlkptqlvapgavvewhetaivpgnltcpin 125


```

OY 139 -----DRELFRPLROECVYTGVTGMPGAVALNSCNDLAGLIRDTSDREFFLE 191
Db 126 nhqpsatyrirktleplqncayvgdivdipgtsvalnsncdglagmlksdneefle 185
OY 192 RGOQKEASGRTHVYVYRREAVOQEAEPDGLH---NEAFGLDLPNLLGLVGDQDTE 248
Db 186 rkgmeeeegrthvvykrsavegapldmskdfhyresdlegldlgtvgnihqnlmetm 245
OY 249 RKRRAKPGSIEVLAVDDSVVRFHKGHNQNVLLMNVDTIYHDESLGHVINITAL 308
Db 246 rrrrhaqendnyevlllyddsvvfthgkehqnylltlnlmeilnydeslghvlnvvl 305
OY 309 VRLVWGYRSLIERGNPSRSLERQVMAHSGOQRODPHAEHNDHVFLVFLROEFGS- 367
Db 306 vrmnllyaksisltergnpsrslenverwasqgrslinhedhaiflrrqdfpbg 365
OY 368 --GYAPVYGMCPRLSCALNHEDGSSAFVIAHETGVLGMEHDOGNGCADETSLGSM 425
Db 366 mgyavpvtgmchprvscldnhdgfsafvvaheghvlgmehdgqgnrcgdetamgsvm 425
OY 426 APLVOAFHFRHMSRCSKLELSRYLPSYDCLLDPEFDPAWQPELPGINSMBOCFD 485
Db 426 aplvvaafhryhmsrscsgelkrylshydclddpfdhwpkldpqlnysmdegcrfd 485
OY 486 FGSGYOTCLAFRTEPCPCOLWCSPHNDNPFCTKKGPRLDTECAPGKWCFCGHCIMKSP 545
Db 486 fgvykmtctafitfdpcqkfwcsphndnpyfctkkgprldtecaagkwcgkymcmwna 545
OY 546 EGYTQODGMSWTKFGSCSRSCGCVARSRSRSCNNPSPAYGPRCLGPMFEYOVCSSE 605
Db 546 ng-qkqdgmswtkfgscsrscgcvarsrsrscnnpspaygprclgpmfeylejntee 604
OY 606 CPGETEDFRAOQCAKRNRYHYHONAKSMWVPEPDDAOKECLLCOSADNDGVFMNQV 665
Db 605 cqkhtedfraqqcgrnshfeygnkthwlpypndpdkrchlycgskedgvaaymqvlv 664
OY 666 HDGRCSIRDPYVCAKECVVGCDEKESKADKCGVCGGNSHCRTRKGTGLGRASK 725
Db 665 hdgthcsydkdyslcvrgecvkvgcdkeigsnkvedkcyvgngshncrtvgtftr 724
OY 726 OAGALKLVQIPAGARHIOIELESPIHRSVKNQVTSFLINPKGKATSTFTFAMGLEW 785
Db 725 klyglkmdlppgarhvlqedeasplhalkqatbhyllngkeakstfildigvew 784
OY 786 EDAVEDAKESLKTSGPPEALAIALPRTGGRSSLAAYKVIHEDLLPLIGSNVLEE 845
Db 785 dnyteddieslhtdgpndpvlvlllpgendtrssltvylilhedsvplinsnnvlee 843
OY 846 MDYEMALKSNAPCSKACGGGIGTGTGCRRRRDHWHVORHLCDHKRRPRIRRCNQHP 905
Db 844 ldtlewalkswsgskpcgqgfytkygcrrckdmkhvhsfceanakprrfrimnige 903
OY 906 CSOPVWTEEMGACSRSCGLGVOTRIGIOLPLINSNGTHVMAKACAGRRPARPCLR 965
Db 904 cltpplwaeeewenctkcgssyqlrtvrcldppldgtnsvnskcmgdrpsrtrpncr 963
OY 966 VPCPAQMRILCAWCSOCSATGEGIQOQVYCRVNTANSLGHCEDRPTVYVCSIPACGNGH 1025
Db 964 vpcpaqmrilqpwseesctvcegtvryqlrag----dhcdgkcpesvraqclppnd-- 1017
OY 1026 QNSTVRADVWELGTPREBQWVPOSGPLHPINIKISSMCAERCTDRSVFCOMEVLDRCST 1085
Db 1018 -----epclgdkslfcqmevlarycsl 1039
OY 1086 PGYHRLCCVSCIRKASG-----PNMGPPPG-----PTSLPPF--STPG- 1121
Db 1040 pygylkccescskrstllpppyllaeaethdvinspsdprslvmpstlvphsetpck 1099
OY 1122 -----SPLPQODPADAAEPGKPRPGSFDHGHGATOLPGALDT-----SSPQTQ 1166
Db 1100 kmslssissvgnpaya-altrpnskpdganlrg--rsaqgagsktvlrvltvpspsptk 1154

```

```

RESULT 3
AAB21254
ID AAB21254 standard; Protein: 1201 AA.
XX
AC AAB21254;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human metalloproteinase KIAA0366.
XX
KW Human: KIAA0366; ADAMTS: metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytosolic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Homo sapiens.
XX
PN W0200053774-A2.
XX
PD 14-SEP-2000.
XX
PE 08-MAR-2000; 2000MO-US06237.
XX
PR 08-MAR-1999; 9905-0264585.
XX
PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
PI Kelnner GS, Clark M, Maki RA;
XX
DR WPI: 2000-594326/56.
XX
DR N-PSDB: AAA95824.
XX
PT Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX
PS Claim 13; Fig 8; 129pp; English.
XX
CC The present sequence is human metalloproteinase KIAA0366. It
CC shows homology to the ADAMTS family of proteins, which contain
CC thrombospondin, integrin and metalloproteinase domains. ADAMTS
CC polypeptides are useful for the manufacture of medicaments for treating
CC conditions associated with neuroinflammation and/or neurodegeneration,
CC such as Alzheimer's disease, Parkinson's disease and stroke. They are
CC also useful for treating conditions associated with cell proliferation,
CC cell migration, inflammation and/or angiogenesis, such as cancer,
CC arthritis and autoimmune diseases. They can be used to treat patients
CC afflicted with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 1201 AA;
XX
Query Match 52.0%; Score 3614; DB 21; Length 1201;
Best Local Similarity 55.5%; Pred. No. 1.5e-248;
Matches 665; Conservative 145; Mismatches 236; Indels 152; Gaps 17;
OY 37 DYGVTVPCTDFRGRFLSHVSGPAAASAGMWDTPTLPRHSHLR-----VARSPLH 91
Db 37 eyelvtpvstnlgrlylshls-----ashkrrsardvsnp-- 73
OY 92 PGGTLPMPGKVGKHSUYFNVTYVCKELHLRLPRNRLVYVPGSSVEMQF----- 138
Db 74 -----eqldntafgkafhlrlkpnqlvargavveheslvpgnltpin 121
OY 139 -----DRELFRPLROECVYTGVTGMPGAVALNSCNDLAGLIRDTSDREFFLE 191
Db 122 nhqpsatyrirktleplqncayvgdivdipgtsvalnsncdglagmlksdneefle 181
OY 192 RGOQKEASGRTHVYVYRREAVOQEAEPDGLH---NEAFGLDLPNLLGLVGDQDTE 248

```

```

Db 182 rgkmeekgrlnhvvykrsavegapldmskdfhyresdlegldltvygnihqnlneim 241
QY 249 RKRHHAKPGSSTIEVLLVVDVSRFHGHKHYONVYLTLMNIYDESLGVHINIAL 308
Db 242 rtrrhagendynlellgvddsvrrfngkhevgnyllltlmvneiyhdeslghvlnvyl 301
QY 309 VRLIMVGYROSLIERGNPSRLSEYVCMAHSQQRDPSPSHAHHDHVFVLTQDPGPS- 367
Db 302 vrmimlgaksisltergnpsrslencrwasqgsrdlnsheshdhaflltrqdlgpag 361
QY 368 --GVAPTGMOHPLRSCALNHDEGFSSAFVIAHETGHVLMGMEHDGOGNGCADSLGSVM 425
Db 362 mgayapvlgmchprscldlnedgissatyaalecgvnlygmenhdggngnrcgdetangsm 421
QY 426 APLVQAFHFRHMSRCSKLELSRPLPSYDCLLDPPDPAMPORPELPGINYSMEDECRFD 485
Db 422 aplvqaafhrhmsrcsgqelkrylhaeydcllddpfdhwpkllpelpglnysmdeqrfid 481
QY 486 FGSGYQTCIARFTEPCQKLMCSHPDNPYCKTKKGPPLDGTGECAPKMGFKGHCIMKSP 545
Db 482 fgvgkmcatafrtdpcqklwscshpdytcktkkppldgtlecaaqkwcgkghcmkna 541
QY 546 EOTVGODGMSWTKFGSCSRSGSVRSRSCNNPSPAYGGRPCUGPMFEYQVCSNSEE 605
Db 542 ng-qkqdgmgswtkfgscsrctgtyrfrtrgcnmpmlnggdcpgvnlfeygqlante 600
QY 606 CPGTYEDEFAOQCAKRNSTYVHONAKHSWPEYRDDDAQKCELICGSADTGDVFNQVY 665
Db 601 cqkhnediraqcgqrnshfeygntklhwljyehpdkkrcchlycqsqkctgdvaaymqylv 660
QY 666 HDGRCFSYRDYSVCAKRECEYVGCDEKVEGSMKADKCGVCGNSGCRVYKGLGKASK 725
Db 661 hdgtrcsyrdydlvscvgeckvycdkdekgcnkvedkcyvgngshncrtckgfttrpr 720
QY 726 QAGALKIVQIPAGARHIOIEALEKSPHRSVYKNOYTSFILNPKGKATSTFTFAMGLEW 785
Db 721 klgyklmctdippgarhvlgedaasphlaikngatghylngrgeekstfidiqevw 780
QY 786 EDVADAKESLKTSGPLPEALAIILPTEGPRSSLAKEYVIEHEDLLPLIGSNVLEEE 845
Db 781 dnyledleslhtdglndpnyvlll-pgendtrssltkyllhnedsvpltnsnvige 839
QY 846 MDTYEMALKSWAPCSKACGGGIGFTKYGCRRRDHVYQRLCDHKRKPFIIRRCQHP 905
Db 840 ldtfewalkswavkpcggfgytkygcrrksdkmwhrfsceankkpkrlrmcniq 899
QY 906 CSQPYWVTEBMCACRSRGKLGVOYRGIOCLPLNSNGTHKVMKAKAGDPEARPCPLR 965
Db 900 cthplvvaeehewckctcgssqyltrvrcqlplldgtrnsnshakymgdprpstprnt 959
QY 966 VPCPAQWRLGAMSOCSATCGEGIGQROYVVCRTNANSLGHCEDRPDIYQVCSLPACGNN 1025
Db 960 vpcpqgwtkgwscsvscvtegrvylcrag----dhcdgkcpesvraclppcnd-- 1013
QY 1026 QNSTVRADVWEIETPEGQWVPSGLPHINKISSMCAAEPECTGPRSVYCOMHEVLDRCYSI 1085
Db 1014 -----epclgdkslitgmeylatrycsi 1035
QY 1086 PGYHRLCVCSTICKKASG-----PNPGPDG---PTSLPFF--STPG- 1121
Db 1036 pgyhrlcvcscstlpppyllaeeahdvinspdlprslvmpscslpylpyshetpak 1095
QY 1122 -----SPLPGPDADAEPPGKPTGSEDHQHRATQPCALDT-----SSBGTQ 1166
Db 1096 kmleslssvsgpnaya-afrpnskpddganlrg--tsaqgsgsktvlrtvpspsptk 1150

```

```

DT 19-MAR-2001 (first entry)
XX
XX DE Bovine metalloprotease.
XX
XX DE Bovine; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
XX cancer therapy; benign tumour; ocular angiogenic disease;
XX rheumatoid arthritis; psoriasis; wound healing; endometriosis;
XX vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
XX scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
XX coronary collateral; cerebral collateral; arteriovenous malformation;
XX ischaemic limb angiogenesis; Osler-Webber syndrome;
XX plaque neovascularisation; telangiectasia; haemophilic joint;
XX angiofibroma; fibromuscular dysplasia; wound granulation;
XX Crohn's disease; atherosclerosis; birth control.
XX
XX OS Bos taurus.
XX
XX PN WO200071577-A1.
XX
XX PD 30-NOV-2000.
XX
XX PF 25-MAY-2000; 2000WO-US14462.
XX
XX PR 25-MAY-1999; 99US-0318208.
XX PR 20-JUL-1999; 99US-0144882.
XX PR 10-AUG-1999; 99US-0147823.
XX PR 13-AUG-1999; 99US-0373658.
XX PR 22-DEC-1999; 99US-0171503.
XX PR 22-FEB-2000; 2000US-0183792.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIR) SMITHKLINE BEECHAM CORP.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PA (IRUE/) IRUELA-ARISPE L.
XX PA (HAST/) HASTINGS G A.
XX PA (RUBEN/) RUBEN S M.
XX PA (JONAK/) JONAK Z L.
XX PA (TRULLI/) TRULLI S H.
XX PA (FORN/) FORNWALD J A.
XX PA (TERR/) TERRETT J A.
XX
XX PI Irueña-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trullli SH;
XX PI Fornwald JA, Terrett JA;
XX
XX DR WPI; 2001-025136/03.
XX
XX PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
XX inhibit angiogenesis in the treatment of disorders such as cancer,
XX rheumatoid arthritis and psoriasis -
XX
XX PS Disclosure; Fig 3; 768pp; English.
XX
XX XX The present invention relates to human METH1 and METH2, (ME for
XX metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
XX The present sequence is bovine metalloprotease, which was used in a
XX sequence homology comparison with the METH proteins. METH can be used
XX for inhibiting angiogenesis in an individual, and for treating cancer,
XX benign tumours, an ocular angiogenic disease, rheumatoid arthritis,
XX psoriasis, delayed wound healing, endometriosis, vasculogenesis,
XX granulations, hypertrophic scars, nonunion fractures, scleroderma,
XX trachoma, vascular adhesions, myocardial angiogenesis, coronary
XX collaterals, cerebral collaterals, arteriovenous malformations, ischaemic
XX limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation,
XX telangiectasia, haemophilic joints, angiofibroma, fibromuscular
XX dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH
XX can also be used in birth control. METH can also be used in diagnostic
XX methods for the prognosis of cancer.
XX
XX SQ Sequence 1203 AA:

```

Query Match 47.8%; Score 3327; DB 22; Length 1203;
 Best Local Similarity 52.2%; Pred. No. 4.2e-228;

Query Match 47.4%, Score 3299.5, DB 19, Length 1211;
 Best Local Similarity 52.6%, Pred. No. 3.8e-226;
 Matches 635; Conservative 157; Mismatches 352; Indels 63; Gaps 20;

1 MAPRLALSTYL-----ELHCACTAGAGSRTPELHLSGKLSDVGVTPCSTDRG 50
 12 lcpallllllllppllppppparlaaadppgprl---ghaaellavprtdagq 68
 51 RELSHVSGPAAASAGSMVVDTPPLPRHSHLRVARSPLHAGTLMGDRGRSHLYRNV 110
 69 lrvshvsa-atsragvarttaap-----vtrpsfpgn--eeepgsh-1fyv 113
 111 TVFGKELRLRLRNRLVVPSSVEKQEDFRLRQECVYRGVTGMGA-AVALS 169
 114 tvfgrelhrlrpnarlvpagatmewqgkyltrvepllgscilygdvagaiaaasvals 173
 170 NCDGLAGLRITDSTDFEFLERGOEKEA-SGRTHVYRRARVAQOEAMEPDG-DLHNEA 227
 174 ncdglagllrimeeeefleplekylaageagrvhvyrrptprlpggpaldtgasl 233
 228 FGLGDLPLNLGLVQDGLGDTERR-RRHARPGSYSLIEVLLVDDSVYRHKHVNLYLT 286
 234 dsldslsrslagyleehansrrrarrhaaddynlellvgddsvyqfngkhhvqkylt 293
 287 LAMIVDEIYHDESLGVHINIALVRLIMGVROSLIIEGNPSRSLIEQCRMAHSGQROD 346
 294 lmiivdeiyhdeslghlnvlyrllllygksmlsliegnpsqlenvcwaaylqgkpd 353
 347 PSHAEHHDVVFLLTRDQFGPS--GYAPYVMCHPLRSCALNHEDGFSSAFVIAHETGV 403
 354 tghdeyhdhailftrtdfgsgmgyparytgmchprscnlhnedgfssafvahnetylv 413
 404 LGMHDSQNGCDEFTSLGVNAPLYQAARHNRHNRSCKLESLRYLPDYDCLLDPEPD 463
 414 lgmhdqgnrcqdevrlygaimaplvqaalhrhwarscqeqlsrylshydcldldpfnh 473
 464 AMQOPRELPGINYSMDQCFDFSGSYQICLAFRTFEPCKOLMCSHPDMPYFCKTKGPP 523
 474 dmpalpqlpjlhysmeqctfdglgymmtatfrtdfcqqlwscndpmyfcktkkppr 533
 524 LDGTECAPGKWCERKHCIMKSPQOTYGGDGMSSWTKFSGSCSGGCVRSRSCNNPS 583
 534 ldgtemcapghckfghciwltpld-llkrtdgswawspfgscsrctgylvfktrtdcndph 592
 584 PAYGRCLEPMFYQVNCNEECRGTEDEPRAOCAKRNRYUYHONAKSHVWEYEDDDA 643
 593 panggrtcslaydfqlscrqdcpsladireegcrtqwdlyfengdaqhncipheindak 652
 644 QKELICOSADTGDVVPMQVNHDRCSYRDPSYVCARCEVPCVGCDEKVGSMKADKC 703
 653 erchlycestrtgevsmkrmvhgtrcsykdaflscvrgdcrkvgdcgylgsskqgdkc 712
 704 GVCGGDNSHCKRTVYGTLAGKASKOAGALKVQIPRAGAHQIDLEAKSRPHNSVYKQVTS 763
 713 gvcggdshckrtvuygtfltrspknhgylkmlfepagahlllqvedadshlnlawnletykr 772
 764 FILMPKCK-EKTSSTFTAMGLEWEDAVEADAKESLKTSGPPEALAIALLPTEGSPRSSL 822
 773 filmeendvdsakftlamyeweyredgetrlqtmgprlghltvlvip--vgdtrvsl 830
 823 AKYVYIHEDLLPLIGSNVNLLEMDYEMALKSWAPSKACGGGTOYTKYKCRRRRHHM 882
 831 tykymihedsin-vddnvlleedsuvyewalkkwspsckpogsgsdtlkygctrrldhkm 889
 883 VQRHLCCHKRRKPRRRRCNQHPCSORPYWVEWEGACSRSGKGVOTRGIOCLPLPSNG 942
 890 vhrfgcaalshpkairracpqcspqrwlvgeewepscqtgrgmgyvtrcqlphdn 949
 943 THKVPARACAGDRPEARPCILRVCPRAOMRLGAMSOCSATCGEGTIOOROVVCFRTNANSL 1002
 950 ttrshahcndarpestrraacsrelcpgtrvragwsgsvtvcgntgetrplclrtadstf 1009
 1003 GHCBSDRPDTYQVCSLPAACGGNHONSTYRADVWELGTPEGGWVQSGPLHINKISSMCA 1062

DB 1010 gigeetpetarctcrlgpcprnlsdpksksyv-----vqwlstrpdpdsplrtkiss--- 1059
 1063 AEPCTGDRSVFCOMEVLDRCSTIPGHRICCVSC-----IKKASGNPDP---DPG 1110
 1060 kphcgdkstfcrmevysrscslpnykllckscnlymltnvgrlepppykhdidvtf 1119
 1111 PTLSPRSTGSPPLPGPDADAAEPPGKPTGSEDHQGRATQLPGLDTSRFGTHP-F 1169
 1120 mtlpvtvmevyrpsptlrvlplnassnatcdhpetnavdepykklngldevgppnl 1179
 1170 APERPIP 1176
 1180 iprrpsp 1186

RESULT 6
 ABG22373
 ID ABG22373 standard; Protein; 1328 AA.
 AC ABG22373;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #22364.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PE 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 DR Dmanac RT, Liu C, Tang YT;
 DR N-PSDB; AAS86560.
 WPI; 2001-639362/73.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

Claim 20; SEQ ID NO 52732; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SO Sequence 1328 AA;

Query Match 47.2%; Score 3279.5; DB 22; Length 1328;
 Best Local Similarity 53.1%; Pred. 1.2e-224;
 Matches 630; Conservative 153; Mismatches 349; Indels 55; Gaps 20;

```

QY 16 CALCTAAGSR--TPELHSGKL---SDYGVTPCSTDFRGRFLSHVYSGPAASAGSMV 70
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 40 crrlllaasrlbqeprevgygprlgbnaerllayprtdagqlvshvsa-atsagrar 98
71 DRRPLRHHSHLRVARSPLHGGTGLMGRGRHSITFNATVFEKELHLRLRNLVLP 130
99 raap-----vtrpsfpggn--eepgsh-lfynvvtigrdhlrlrparlvap 144
131 GSSVEMQDFRELFROPLROECVYTGVTGMPGA-AVAISNCODLACLRDSTDFIEP 189
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 145 galtemwqgkqtrvepllgsclyvgdvaglaassvalncdglaglimeeeffiep 204
QY 190 LERGOQKEA-SGRTHVYVREAVQOEAEPDQ-DLHNEAFGLDLPNLLGLVQDLDGT 247
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 205 lekylaqaqaegrvhvvyrrptsppljggpqaldtgasldslsralyglehans 264
QY 248 ERK-RRNAKPGSYSEVLVVDSDVFRHGRKEHQNVLTLMNVDEIYHDESLGVHIN 306
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 265 rrrarhaaddynlevllgydvsvvfngkhenykyllclmlvmeiydeslghahlv 324
QY 307 ALVRLIMGYROSLLTERGNPSSLEQVCRMAHSQQRDPSSAENHDHVFPLRQDFGP 366
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 325 vlvrlilysygsmllelgnpsqslenvcrwaylqgkptdgydeyhdaifltqdfg 384
QY 367 S---GIAPVTGMCPLRSCALNHEDGFSSAFVIAHENGHLGHEHGOGNCADETSLS 423
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 385 sgmgypapvymchprvscclnhdgfsaavahghvlgmehdgqgrcdevrlls 444
QY 424 VMADPLVQAARFRHFWRSCKSLERLYLPSYDCLLDPPFAPMPQPELPGINSMDEQR 483
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 445 imaplvgaaafnrfmwsrscgelsrylshysdclldpfaidwpaqlpqlhlysmegcr 504
QY 484 FDPGSGYOTCIARFTEPCQIOWCSHPDNPFEKTKKGPPLDGTBCAPKWCFCGKCIWK 543
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 505 fdrlgyymcalftrtdpcqklwscnmpdnyfcktkkppldgtumcapgkchckfncial 564
QY 544 SPEQTYQDGGWSSMTKFGSGSRSCGAGVRSRSCNNPSPAGGRCPLGPMFEYQVCS 603
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 565 tpd-llkrdsgwagwspfgscstcgtygkftrrgcdmhpangrtscglaaydtqlcsr 623
QY 604 EECGTIEDFRAOQCAKRNYSYVHONAKHSWYPERDDAQKCELLQASADTGDVVFMO 663
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 624 qdcpsdlsadfreecgrwdlyfehgdahhlwlphehrdaekchlyceeregevvsmkr 683
QY 664 VHDGTCSYRDYSYVCARGECYPCGCDKEVSGMKADDCGVCGGDNHCHRTVKGTLGKA 723
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 684 mwtdgltcsykdaflscvqdcvkycgdvlgsskgedkcyvgcdnsckvkvkfltrs 743
QY 724 SKQAGALKIQLVAPAGRHQIQLALEKSPHRSYVKNQVTSFLINPKGK-ENASRTFTMG 782
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 744 pkhgyikmelpagarrhlllgevdatslhavknlctgkflleendvdasskclfiag 803
QY 783 LEWEDAVEDEAKESLTSGLPEALIALPTEGGFRSSLAYKYVIHEDLPLIGSNVL 842
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 804 veweyrddedrelqlmglhgtlvtlvlp--vgdtrvsltykynlhedsln-vddnnvl 860
QY 843 LEEMDTYEMALKSWAPSCACGAGIOFTYKGRRRDHMMQRHICDKRKPRRRRCN 902
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 861 eedsvvyewalkkwspcsrpgcgsgqfctkygrrtrldhmwvtrgfaalskpkalraen 920
QY 903 QHPCSQPVWTEEMWAGSCSGKGLGVQTRGICLPLNSGTHKVPAPAKAGDPEARRP 962
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 921 pgecsqpvwvtgewepscgtcgtgmgyvsvrciqlndnttrshahcndapestra 980

```

```

QY 963 CLRVPCPAQWRLGAMSOCATGEGEIQOQRYVCFERNANSLGHCGBRDPVQVCSLPAAG 1022
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 981 csrelcprgrwvrgpwsqsvlcognlgerpvtcradastglqeeqpeartcrlrpgcp 1040
QY 1023 GHNQSTVRADYVWELGTREGQWVPOSGPLHPINKISSMCAEPCTGDSRYFCOMEVLDY 1082
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 1041 rnsdpskksy-----vqlsrpdpdpplrklss---kghcgqskslfctmewlsry 1090
QY 1083 GSIPGYNRLCCVSC-----IKKASPNRGP----DPGRTSLPSTGSPLPGRQDP 1130
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 1091 csigqykrllqkscllynpnrvqgriepgpkhndldvimpdlpyvtvamevrpsstp 1150
QY 1131 ADAAEPKPKPTGSEDPHQGRATOLPGALDTSSTPGTQH-PAPERPLP 1176
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 1151 levpnasslnatedhpetnavdepyklhngldevqpnllirppsp 1197

```

RESULT 7

AAW47030 standard; Protein: 1206 AA.

AAW47030;

06-JUL-1998 (first entry)

Bovine N-proteinase.

N-proteinase; cattle; collagen; antibody; rheumatoid arthritis;

fibrosis; Ehlers-Danlos disease; diagnosis; therapy.

Homo sapiens.

```

FT Key Location/Qualifiers
FH MISC-difference 142
FT /note= "encoded by CCC"
FT MISC-difference 207
FT /note= "encoded by CCG"
FT MISC-difference 464
FT /note= "encoded by CCC"
FT MISC-difference 545
FT /note= "encoded by ATC"
FT MISC-difference 549
FT /note= "encoded by CCG"
FT MISC-difference 810
FT /note= "encoded by ATC"
FT MISC-difference 980
FT /note= "no equivalent codon in nucleotide sequence"

```

WO9800555-A1.

08-JAN-1998.

03-JUL-1997; 97WO-US12427.

02-JUL-1997; 97US-0886333.

03-JUL-1996; 96US-0021203.

(COLI/) COLIGE A.
 (LAP/) LAPIERE C.
 (PROC/) PROCKOP D J.

Collage A, Lapierre C, Prockop DJ.

WPI, 1998-086980/08.

N-PSDB; AAV06594.

Poly-nucleotide sequence encoding human N-proteinase - used to
 produce mature collagen in vitro and antibodies to treat fibrosis
 and rheumatoid arthritis

Disclosure; Fig 5; 49pp; English.

This polypeptide comprises human N-proteinase, the enzyme responsible for cleaving N-propeptide from procollagen to produce mature collagen. The amino acid sequence was deduced from a cDNA sequence (see AAV06594) derived from bovine cDNA clones. Bovine N-proteinases can be produced using a claimed method in which a host cell is transformed or transfected with an N-proteinase cDNA, and the N-proteinase is isolated from the medium. The N-proteinase can be used for the production of mature collagen *in vitro* and for the production of antibodies which may be used for diagnosis and therapy of diseases including fibrosis and rheumatoid arthritis. The N-proteinase may also be administered to treat a disease resulting from insufficient production of N-proteinase, such as Ehlers-Danlos disease. Human N-proteinases (see AAW47028-29) are also disclosed.

Sequence 1206 AA:

Query Match 47.0%; Score 3270.5; DB 19; Length 1206;
 1st Local Similarity 51.7%; Pred. No. 4.4e-224;
 Matches 636; Conservative 160; Mismatches 356; Indels 79; Gaps 21;

7 LLSLLEHLCALCTAAGSRPELHLSGLSDYGVTPCSTDFGRFELSHVSGPAAASAG 66
 17 LLLLLLPadarlaaaadpppgpghgaerllavpvtcdagrlvshvsa-atapag 75
 67 SMVVDTPPTLRHSSHLRVARSPHGLTLPGRVGRHSLYFVYVFKELHLRLPNRR 126
 76 ---Vrttraapaqipgls-ggseedpgg-----rlfynvtvgrdlhrlrpnar 121
 127 LVVGGSSVENQEDRELFRQPLROECVYTGVTGM-PGAVAISNCGGLRLRTSDTF 185
 122 lvapgatvewgsegatrvellgtcltygdvaglaessvalncodlagllrmeecf 181
 186 FIEPLERGOOEKKA-SGRTHVYVYRRAVVOQEMAEPPD-----DLHNAFGLGDLPNLLGL 239
 182 flieplekylakeeggrvvhvhrftsr--ppplgspaldgtisadsidsisralgy 239
 240 VGDLGDTERR-RRAKFGSISTEVLVYDSDVYRFHGEKVQVYVLTINIVDEIYHDE 298
 240 leevnsarrmrhaaddnyllgyddsvvqfngtehvqgyltllmlyelxhde 299
 299 SLGVHINIALVRLVWGRQSLIERGNPSRLQVCRMAHSQORODPSHAHHHNVF 358
 300 slgahinivvrllyllygkmslilegnpsqlenwcrway-lgkprcdthdeydhalf 359
 359 LTRQDFGS---GYAPVTGMCHPLRSCALNHEDGSSAFVLAHETGHVYLGHENHOGNGC 415
 360 ltrqdfgsgmggyapvymchpvrscclnhedgfssafvahalghvlgmehdggnrc 419
 416 ADESLGSLVMAPIVOAAFRFHMHSKSLRLSPYDCLLDPPFPAMPQPELPGIN 475
 420 gdevrlgsmaplvaafhrfwrscsqelsrlyhsydcrlrdffcthdwlpalqlp9lh 479
 476 YSMDOCRDFEGSGYOTCLAFRTPEPCOLWCSPHDPNPFCKTKKPPLDSTBCAPKGC 535
 480 ysmecrfdfglgymctaftrfdpcckqlwoshpndpyfcktkkppldltmcapghc 539
 536 FKGRHCTKSPEDTYGDDGWSMTKEGSCRSYCGGVSRSRSCNPNPAPAGRPCLGPM 595
 540 fkhctwtlfd-llkrdgnwagwspfgscstcgvvfrftrqcanpnapagrtcs9la 598
 596 FEYQYCNSEPCGTEDPRAQCAKRNSTYHQNAKHSWVPEPDDAQKCELICQASDT 655
 599 ydfqcnscdcpdaladfrteegctqwdlyfchgdqhlwphrehdakerechlycesket 658
 656 GGVTFMNQVYHDGTRCSTRDPYVCARGECYVCGDKKEYGSKADKCGVGGDNHSHRT 715
 659 gevtmkmrhvhdgtrcstrdpysdafsclcvrgdckvsgcdgylysgskdgcyvggdshskv 718
 716 VKGTLGAKKQKALKIVQIPAGARHIQIEALEKSPHRSVSNQVTGSEILNPKCK-DAT 774

719 vkgtfrspkkllykmtelpagarhlllgadtstshlavknuletykllneendv9n 778
 775 SRPTAMGLEMEDAVEAKESLKTSGPLPEALIALPPTBEGGRSLAKKYIHEDLP 834
 779 sktflamgveweydedygrelqltmprhgtltvlp--egdarisltykmlhedsln 836
 835 LIGSNVLEMDYHVMALJKSMAPCSKACGGGIGFTFYGGCRRRRDHMMVORHLCHHKRP 894
 837 -vdannvledsdvgyewalkkwpsscpkgsgsfkygcrtridhkmvnhgfdsvskp 895
 895 KPIRRRCNOHPCSQPVWTEEMGACSRGKLGVTQRIQCLPLNSGTIKVMARKACAG 954
 896 kairrncnpgcesqgvwvltgewepscgrltgmqvsvrvcvqplhnnltvshkncd 955
 955 DRPEARRPCILVPCPAQORLAMS-OCSATGCEGIGQROVQCFRMANSLGHCEDRDPDV 1013
 956 arpegrracnrcilcpgrtragswsgcsvtcngntgterpylctrcadsfygcrepeta 1015
 1014 QVCSLPACGNGHONSTVRADYWELOTPBQWVPOSGPLHPINKISSMCAEPCTGDSVF 1073
 1016 ricrlgprpntsdpskksy-----vqlsrpdpnspvqetss--kgrcgdksvf 1065
 1074 COMEVLDRKCSIPGTHRLCCVSC-----IKKASGPNPGPPGPTSLPFPSPGSP 1124
 1066 crmewlsrscsipygnklcckscnphdnltdvddraepsskhnideelmltlyvltvm 1125
 1125 ---PGROPADAAEPKPGTGESEDHQHGRRATQLPGALDTSRGTQHPAPPTPIPGASWS 1181
 1126 ewgpppglpvleplntstsnatdhpctnnavdpykl-----pgledevqpnllp----- 1176
 1182 ISPTTGGLPWCGWOTPTTPVPEDKQPGEDL 1212
 1177 -----rrpspyektlnqrigel 1193

RESULT 8
 AAT53900
 ID AAY53900 standard; Protein; 969 AA.
 XX
 AAY53900;
 XX
 13-MAR-2000 (first entry)
 XX
 DE Amino acid sequence of a bovine PINP protein.
 XX
 KW GON-1; metalloproteinase; cell migration; modulator; PINP;
 KW metalloproteinase domain; thrombospondin domain; abnormal cell migration;
 KW organ shaping; sterility; cancer metastasis.
 XX
 OS Bos sp.
 XX
 PN WO961656-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 28-MAY-1999; 99WO-US11918.
 XX
 PR 29-MAY-1998; 98US-0087170.
 PR 13-APR-1999; 99US-0129023.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Kimble JE, Bieleloch RH;
 XX
 DR WPI; 2000-072633/06.
 XX
 PT Identifying modulators of proteins containing metalloproteinase and
 PT thrombospondin domains, potentially useful for controlling cell
 PT migration and organ shaping -
 XX
 PS Disclosure; Fig 1C; 60pp; English.
 XX
 CC The present sequence represents a bovine PINP protein. PINP is

CC a metalloproteinase. The specification describes another related
 CC metalloproteinase, a C. elegans GON-1 protein, that lacks a transmembrane
 CC domain and possesses a predicted metalloproteinase domain between residues
 CC 269-456. In C. elegans hermaphrodites, GON-1 is required for migration of
 CC two distal tip cells to produce elongated tubes, whereas in males, GON-1
 CC is required for migration of a single linker cell to produce a single
 CC elongated tube. The protein is used in the method of the invention. The
 CC specification describes a method for identifying a modulator of a
 CC protein that contains a metalloproteinase domain and a thrombospondin
 CC domain. The method comprises treating a target organism, having a
 CC developing gonadal cell that is responsive to the protein, with a test
 CC compound, and determining any change in migration or shape of the cell
 CC attributable to the test compound. The compounds identified are
 CC potential therapeutic modulators of abnormal cell migration and organ
 CC shaping, e.g. for rendering animals (specifically nematodes) sterile
 CC and for inhibiting cancer metastases.

Sequence 969 AA:

Query Match 42.6%; Score 2962.5; DB 21; Length 969;
 Best Local Similarity 57.4%; Pred. No. 2,9e+202;
 Matches 559; Conservative 123; Mismatches 259; Indels 33; Gaps 14;

QY 7 LLSYLLPRLCALCTAAGSRTPELHLSGLSDYGVTPCSTDFRGRFLSHVSGPAAASAG 66
 Db 17 LLLLLLPpadarlaaaadppgpgbgaerilavpyrtdagrlvshvsa-atapag 75
 QY 67 SMVVDPTPLPRHSHLRVARSPLHGGTLMGCRGRSLRYNTVYFCKELHLRIENRR 126
 Db 76 ---vrrraapaglpigls-gseedpgg-----rlfynvtfgrdlhlrlpar 121
 QY 127 LVVPGSSVMEQEDRELRLROECVYTGVTGM-PCAATAISNCDCIAGIRFSDTF 185
 Db 122 lvpagatvewgesgslrveplglcylvgdvaqlaessvalncdglalimeeef 181
 QY 186 FLEPLERGOQKEA-SGRHVVYRKREAVOQEAEPDG-----DLHNEAFGLDPLNLTGL 239
 Db 182 flleplekylaaakeagqrhvhyhprltsr--pprl99pqaIdtqisadsidsraIga 239
 QY 240 VEDDLDTDRK-RRIAKPGSYSTEVLVVDSSVYRFGKEHVNQVLTIMTYEITHDE 298
 Db 240 leevnssrrmrtaadddynlellygvdsvvqfthetvqkylltlimiveIynde 299
 QY 299 SLGVHINALVLIWGVTRQSLIERGNPSRLSEQVCRMAHSQRODPSSHAEHHDHVF 358
 Db 300 slgahlnvrlrlllsyqksmllelgnpsqlenverwaylqkprcdtdheydhaf 359
 QY 359 LTRDFGFS---GYAPVGMCHPLRSCLNHDGFSFVLAHETGHVLEMEHDGNGC 415
 Db 360 ltrdfgspgmgyavpvgmchpvrscldnhdgfssafvahetghvlgmehdggnrc 419
 QY 416 ADESLIGSVMAPLVOAARFHRFHSKSLERLRLPSYDCLDDPPRPAMPQPELPGIN 475
 Db 420 gdevlrglmaplvgaefhrfhwercsqelsrlylshydcldrpfhndwpaIpgIh 479
 QY 476 YSMDOGRFDFSGGQTLARTRPECKQIWCSPNDNPFCKTKKGPDLGTCAAPKWC 535
 Db 480 ysmegctidfglymmctatrltdpckqlweshpndpcklkkppldgtmcapgknc 539
 QY 536 FKSHCNKSPQOTGODGGSMTKFGSCRSRGGGRSSRSCNNNSPAYGGRPLCGPM 555
 Db 540 fkgshwtltpd-llkrdgnwgaawpfsgscstcgvykfltrcdpdpbanagrtscgla 598
 QY 596 FEYOVNNEECPGTYEDFRAOQCAKRNSSYYVHONAKSHWPPEPDDAOKCEITICOSAP 655
 Db 599 ydficnscdpcdaldafreeqcrqwdlyfehgaqhwlphehtrakercchlyesket 658
 QY 656 GDVFMNVNVADGTRCSRDPSYVARGCEVPVCGCKEVSMAADCKCGGSDNSHCT 715
 Db 659 gevsvmkrmvndgtcsykdafslcvrgdcrkyvgcdgvisqskedkvcygcgdnshckv 718
 QY 716 VKGTGLKASKAGAIKLVOIPAGARHIOLEALEKSPHRSVYKQVYGSFILNKGK-EAT 774

Db 719 vkgtfsspkllgikmfepagarrlllqeadtshlavlknletgkllneendvgn 778
 QY 775 SRTTAGLMEDEAVEDEAKESLTKSGPLPEALIIALPTREGPSRLAYKVIHEDLLP 834
 Db 779 sktliangvreydeagretlqtmplngltvlvlp--eddaristlykymhdedln 836
 QY 835 LIGSNVLLDEMDTYEMALKSMAPCSKACGGIOFTKYGCRRRRDHMYORHLCDHKRRP 894
 Db 837 -vddnvnleddsvyewalkkwspckpccgsgqftkygcrirrlbhkwhrfcdsvskp 895
 QY 895 KPIRRRQNHPCQPPWVYIEMWACGRSGCKIGVOTRIGIQLPLPSNGTHKWPAAKACAG 954
 Db 896 kairltcnpgcsqpvvltgvepcscrgtgmgyvsvrcvqlhnttsvhtkncnd 955
 QY 955 DRPEARPPCLRVPC 968
 Db 956 arpegrtracnrelc 969

RESULT 9
 AAM47029
 ID AAM47029 standard; Protein; 566 AA.
 XX
 AC AAM47029;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Human N-proteinase (70 kDa short form).
 XX
 KW N-proteinase; human; collagen; antibody; rheumatoid arthritis;
 KM fibrosis; Ehlers-Danlos disease; diagnosis; therapy.
 XX
 OS Homo sapiens.
 XX
 PN W09800555-A1.
 XX
 PD 08-JAN-1998.
 XX
 PF 03-JUL-1997; 97WO-US12427.
 XX
 PR 02-JUL-1997; 97US-0886333.
 PR 03-JUL-1996; 96US-0021203.
 XX
 PA (COL1A) COL1A A.
 PA (LAP1) LAP1ERE C.
 PA (PROC) PROCKOP D J.
 PI Collige A, Lapierre C, Prockop DJ;
 DR WPI: 1998-086980/08.
 DR N-PSDB: AAV06593.
 XX
 PT Polynucleotide sequence encoding human N-proteinase - used to
 PT produce mature collagen in vitro and antibodies to treat fibrosis
 PT and rheumatoid arthritis
 PS Disclosure: Fig 2B; 49pp; English.
 CC
 CC This polypeptide comprises the 70 kDa short form of human
 CC N-proteinase, the enzyme responsible for cleaving N-propeptide from
 CC procollagen to produce mature collagen. The amino acid sequence
 CC was deduced from a cDNA sequence (see AAV06593) derived from human
 CC skin fibroblast cDNA clones. The 130 kDa long form (see AAM47028) of
 CC human N-proteinase was also identified. Human N-proteinases
 CC can be produced using a claimed method in which a host cell is
 CC transformed or transfected with an N-proteinase polynucleotide
 CC sequence, cultured in an appropriate culture medium, and the
 CC N-proteinase is isolated from the medium. The N-proteinase can be
 CC used for the production of mature collagen in vitro and for the
 CC production of antibodies which may be used for diagnosis and
 CC therapy of diseases including fibrosis and rheumatoid arthritis.
 CC The N-proteinase may also be administered to treat a disease

AA 25-JUN-1999; 99JP-0180973.
PF
XX

Db 602 vqgshfda-mlykqglhtwpvv--ndvnpceclhcrpaneyfaeklrdavvdgltcpcyqr

675 DPYSVCARGECPVPGCDKEVGSMDKDDKGVCGGDNSHCRFVKGTIGKASKOAGALPLKVO

```

Db 659 asrdicngicKnygcdfeidsamedrcychnsgstchtvsftfeea eglyvwdvql 717
QY 735 IPAGARHIOIFALEKSPHSRVKNOVTGSEFLNPKGEATSRTFTAMGLEWEDAVEDAKE 794
Db 718 ipagareirIgeveaanflalrisedekyflngwrlgwndygyagtlftyarrgnwe 777
QY 795 SLKTSGPPLPEAIALALPRTGGRSSLAKEYIHEDLPLIGSNNVLEEMDTYEMALK 854
Db 778 nltpspkpevwlqll-fgesnp--gvhyeylhtre---agshdev--ppvfwshyg 828
QY 855 SWAPCSACGGGIGFTYKGGRRRRHHMYORHLCDHKRKPPIRRRONHPCSQPV-WWT 913
Db 829 pwtctvlgcgrvgrnyvclerqagrvdeeh-cdplgrpddqgrkcsqpc--parwa 885
QY 914 EEWGACSRSCCKLGVOIRGICQLPLPLSNGTHKVMAPAKACAG-DRPEARRRCLR-VPCPAQ 971
Db 886 gewqlcsscpgpqlsrravlcitrsygldegsaleppacehlprppteletcmrhyvcpat 945
QY 972 WRIGAMSQCSATCGEGIOQRQVVCRTNANSLGHCEGDRPDVTQVCSLPACGGNHQNSTVR 1031
Db 946 wavgwsgcsvtcgegtqrtnvlic-tndtyvpccdeaqpsasevtcsIplcr----- 995
QY 1032 ADVWELET--PEGQWVPOSG-----PLHPINKISSMCAEPCTGDRSVFCOMEVL 1079
Db 996 ---wplqtlqpegsgsgssshelIneadflphlprpspasspkpqlmgnaleeaepel 1052
QY 1080 DRYCSIPG-----YHRLCCVSCIKKAS-GPMRPGD-----PGTSLPPSTPG 1121
Db 1053 d-----lpgvfvddfydydnfinfhedlisygsseepdlagtgdrtpphshpaapstg 1108
QY 1122 SPLRGQDPADAAE---PQKPTGSEDHOGH---ATOLPG-ALDTSPPGTQHPF-AP 1171
Db 1109 spvpateppaakeegvlgpwsps-pwpsqagrspppsseqtpgnpllnflpeedtpigap 1167
QY 1172 EFPPIPGASWS-----ISPTTP-----GCLPMGWOTPTPVPEDKGGPGED 1211
Db 1168 dIglpalswprvstdglqtpacpbesqndfpygkdsqslppwrtdrtnevfkdeeep--- 1224
QY 1212 LRHPGTSLPADLPGRPE---PCHPTGT 1236
Db 1225 ---kyrgap-hlprpstlpplspvgs 1248

```

Fetch completed: August 28, 2002, 20:10:28
Time: 1100 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 22:35:24 ; Search time 61.11 Seconds

(without alignments)
361,697 Million cell updates/sec

Title: US-09-938-330-22_COPY_1040_1249
Perfect score: 1206
Sequence: 1 PEGQWVPGSGPLHPINKISS.....PFGTFLICVPRDSQLRGHT 210

Indexing table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Database: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID55/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID55/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	84.6	1048	22	AA85695*
2	217.5	18.0	1211	19	AAW47028*
3	211.5	17.5	1203	22	AA850004*
4	211.5	17.5	1206	19	AAW47030*
5	204	16.9	1201	21	AA821254*
6	204	16.9	1205	22	AA873550*
7	188.5	15.6	1328	22	AA823373*
8	175	14.5	762	16	AA880327
9	175	14.5	762	16	AA880324
10	175	14.5	762	16	AAW49713
11	175	14.5	762	19	AAW49715

12	175	14.5	762	19	AAW57666	Collagen-like poly
13	175	14.5	762	19	AAW57668	Collagen-like poly
14	175	14.5	762	19	AAW53535	Amino acid sequenc
15	159.5	13.2	635	22	AA828860	C glutamicum prote
16	158	13.1	682	16	AA80330	Protein polymer ad
17	158	13.1	682	19	AAW49717	Collagen-like poly
18	158	13.1	682	19	AAW57670	Collagen-like poly
19	154	12.8	417	18	AAW26354	Collagen-like poly
20	154	12.8	417	18	AAW26354	Collagen-like poly
21	154	12.8	417	21	AAW53531	Collagen-like poly
22	154	12.8	417	21	AAW53531	Collagen-like poly
23	154	12.8	837	16	AAW82821	Amino acid sequenc
24	154	12.8	837	16	AAW80320	CLP 3.7 amino acid
25	154	12.8	837	19	AAW26353	Collagen-like prot
26	154	12.8	837	19	AAW57659	Collagen-like poly
27	154	12.8	837	21	AAW53530	Collagen-like poly
28	152.5	12.6	829	16	AAW80334	CLP 3.7 amino acid
29	152.5	12.6	829	19	AAW57673	Protein polymer ad
30	152.5	12.6	829	19	AAW57673	Protein polymer ad
31	151	12.5	448	21	AAW53201	Collagen-like poly
32	150	12.4	330	19	AAW57645	Macaca mulatta rha
33	150	12.4	408	17	AAW07539	Collagen-like poly
34	150	12.4	408	22	AAW72737	Collagen-like poly
35	150	12.4	408	22	AAW64007	Repetitive protein
36	149.5	12.4	1012	20	AAW17406	CLP protein sequen
37	149.5	12.4	1196	13	AAW28916	Human atrophin-1 r
38	148	12.3	1664	19	AAW43106	Type III procollag
39	145.5	12.1	1255	19	AAW52249	C. thermocellum ol
40	145.5	12.1	1255	20	AAW6733	Mouse Rho target p
41	145	12.0	623	18	AAW12843	Mouse mdia Rho tar
42	145	12.0	626	18	AAW12842	Pro-alpha (I) (I)
43	145	12.0	1180	22	AAW65175	Truncated pro-alph
44	145	12.0	1466	22	AAW50291	Drosophila melanog
45	145	12.0	1469	22	AAW51591	Collagen type III
						Novel human diagno

ALIGNMENTS

RESULT 1	AA85695 standard; Protein; 1048 AA.
ID	AA85695;
XX	
AC	AA85695;
XX	
DT	29-OCT-2001 (first entry)
XX	
DE	Larvax carcinoma associated protein-1 (LarCAP-1) polypeptide.
XX	
KW	Larvax carcinoma associated protein-1; LarCAP-1; carcinoma; cytostatic;
KW	antimetastatic; antiarthritic; osteopathic; immunosuppressive; human;
KW	immunostimulant; cerebroprotective; vasotropic; antiinflammatory;
KW	dermatological; cardiact; vaccine.
XX	
OS	Homo sapiens.
XX	
PN	WO200159133-A1.
XX	
PD	16-AUG-2001.
XX	
PE	12-FEB-2001; 2001WO-EP01525.
XX	
PR	14-FEB-2000; 2000EP-0102955.
XX	
PA	(MERE) MERCK PATENT GMBH.
XX	
PI	Duecker K, Hentsch B, Hoheisel J, Frohme M;
XX	
DR	WPI; 2001-483569/52.
XX	
PT	N-PSDB; AAH47048.
PT	Larvax carcinoma associated protein-1 polypeptide for the treatment of
PT	carcinomas, metastasis, arthritis, osteoporosis, immune disorders,

PT stroke, ischemia, autoimmune disease, angiogenesis, skin disorders and
PT organ malfunctions -

PS Claim 1; Page 33-39; 47pp; English.

XX This represents a human larynx carcinoma associated protein-1 (LarCAP-1)
CC polypeptide. The LarCAP-1 polypeptide can be expressed by standard
CC recombinant methodology. The LarCAP-1 polypeptide and polynucleotide are
CC useful in diagnostic assays and for the treatment of carcinomas,
CC metastasis, arthritis, osteoporosis, immune disorders, stroke, ischemia,
CC autoimmune disease, angiogenesis, skin disorders and organ malfunctions
CC especially heart hypertrophy.

XX Sequence 1048 AA;

Query Match 84.6%; Score 1020; DB 22; Length 1048;
Best Local Similarity 98.4%; Pred. No. 7,4e-64;
Matches 179; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 PEGQWVPSGRLPIRINKISSTPECTGDRSVFCOMEVLDRIYCSIPGYHRLCCVSCIRKASG 60
|||||
865 PEGQWVPSGRLPIRINKISSTPECTGDRSVFCOMEVLDRIYCSIPGYHRLCCVSCIRKASG 924

QY 61 PNEGPDGPTSLPFPSTPGSPLEGPQDPADAAEPKPTGSEDHQGRATQLEGALDTSS 120
|||||

Db 925 PNPQPDGPTSLPFPSTPGSPLEGPQDPADAAEPKPTGSEDHQGRATQLEGALDTSS 984
|||||

QY 121 PGTHPAPERTPIRGASMSISPTPGGLPMGWTQTPPVEDKGGPCEDLRHGTSLPAD 180
|||||

Db 985 PGTNPAPERTPIRGASMSISPTPGGLPMGWTQTPPVEDKGGPCEDLRHGTSLPAD 1044
|||||

QY 181 LP 182

Db 1045 sp 1046

RESULT 2

AAM47028
ID AAM47028 standard; Protein; 1211 AA.

XX AAM47028;

DT 06-JUL-1998 (first entry)

DE Human N-proteinase (130 kDa long form).

XX N-proteinase; human; collagen; antibody; rheumatoid arthritis;
KW fibrosis; Ehlers-Danlos disease; diagnosis; therapy.

XX Homo sapiens.

XX WO9800555-A1.

XX 08-JAN-1998.

XX 03-JUL-1997; 97WO-US12427.

XX 02-JUL-1997; 97US-0886333.

XX 03-JUL-1996; 96US-0021203.

XX (COLI/) COLIGE A.

XX (LAPI/) LAPIERE C.

XX (PROC/) PROCKOP D J.

XX Collige A, Lapiere C, Prockop DJ;

XX MPI; 1998-086980/08.

XX N-PSDB; AAV06592.

PT polynucleotide sequence encoding human N-proteinase - used to
PT produce mature collagen in vitro and antibodies to treat fibrosis
PT and rheumatoid arthritis

XX Disclosure; Fig 1D; 49pp; English.

PS This polypeptide comprises the 130 kDa long form of human
XX N-proteinase, the enzyme responsible for cleaving N-propeptide from
CC procollagen to produce mature collagen. The amino acid sequence
CC was deduced from a cDNA sequence (see AAV06592) derived from human
CC skin fibroblast cDNA clones. The 70 kDa short form (see AAM47029) of
CC human N-proteinase was also identified. Human N-proteinases
CC can be produced using a claimed method in which a host cell is
CC transformed or transfected with an N-proteinase polynucleotide
CC sequence, cultured in an appropriate culture medium, and the
CC N-proteinase is isolated from the medium. The N-proteinase can be
CC used for the production of mature collagen in vitro and for the
CC production of antibodies which may be used for diagnosis and
CC therapy of diseases including fibrosis and rheumatoid arthritis.
CC The N-proteinase may also be administered to treat a disease
CC resulting from insufficient production of N-proteinase, such as
CC Ehlers-Danlos disease.

XX Sequence 1211 AA;

Query Match 18.0%; Score 217.5; DB 19; Length 1211;
Best Local Similarity 35.4%; Pred. No. 2.3e-07;
Matches 51; Conservative 14; Mismatches 66; Indels 13; Gaps 3;

QY 4 QWVPSGRLPIRINKISSTPECTGDRSVFCOMEVLDRIYCSIPGYHRLCCVSCIRKASG 55
|||||

Db 1043 QWVPSGRLPIRINKISSTPECTGDRSVFCOMEVLDRIYCSIPGYHRLCCVSCIRKASG 1102
|||||

QY 56 KRASGPNPGR---DPGPTSLPFPSTPGSPLEGPQDPADAAEPKPTGSEDHQGRATQ 111
|||||

Db 1103 EGRIEPPGKHNIDIVFMTPLPVIVAMEVVRPSPTPLEVINAESTNATEDHPETNAV 1162
|||||

QY 112 LFGALDTSSPGTQHP-FAPETPIP 134
|||||

Db 1163 EPYKIHGLDEDEVGPNNIIPRPS 1186
|||||

RESULT 3

AAB50004
ID AAB50004 standard; Protein; 1203 AA.

XX AAB50004;

DT 19-MAR-2001 (first entry)

DE Bovine metalloproteinase.

XX Bovine; METH; metalloproteinase; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischemic limb angiogenesis; Osler-Weber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiodioma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.

XX Bos taurus.

XX WO200071577-A1.

XX 30-NOV-2000.

XX 25-MAY-2000; 2000WO-US14462.

XX 25-MAY-1999; 99US-0318208.

XX 20-JUL-1999; 99US-0144882.

XX 10-AUG-1999; 99US-0147823.

XX 13-AUG-1999; 99US-0373658.

```

PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMIK-) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PA (JONA/) JONAK 2 L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
XX
IRUELA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
FORNWALD JA, Terrett JA;
WPI: 2001-025136/03.
XX
PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX
PS Disclosure: Fig 3; 768bp; English.
XX
CC The present invention relates to human METH1 and METH2, (ME for
CC metalloproteinase and TH for thrombospondin; see AAB50002 and AAB50003).
CC The present sequence is bovine metalloproteinase, which was used in a
CC sequence homology comparison with the METH proteins. METH can be used
CC for inhibiting angiogenesis in an individual, and for treating cancer,
CC benign tumors, an ocular angiogenic disease, rheumatoid arthritis,
CC psoriasis, delayed wound healing, endometriosis, vasculogenesis,
CC granulations, hypertrophic scars, nonunion fractures, scleroderma,
CC trachoma, vascular adhesions, myocardial angiogenesis, coronary
CC collaterals, cerebral laterals, arteriovenous malformations, ischemic
CC limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation,
CC telangiectasia, haemophilic joints, angiodiroma, fibromuscular
CC dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH
CC can also be used in birth control. METH can also be used in diagnostic
CC methods for the prognosis of cancer.
XX
SQ Sequence 1203 AA;

```

```

Query Match 17.5%; Score 211.5; DB 22; Length 1203;
Best Local Similarity 29.1%; Pred. No. 6e-07;
Matches 52; Conservative 24; Mismatches 68; Indels 35; Gaps 4;
4 QWVPGSGPLHPINKISTEPCGDSVFCQMEVLDRCISIPGYHRLCCVSC----- 54
DB 1035 qwlstrpdpnspvqelskgrcgdksvfcrmevlsrscsipygnklckscnphnltlv 1094
QY 55 IKKASGPNPGDPGPTSLPFPSTPGSPPL---PGQDPADAAEPPEKPTGSGEDHOGHGRATQ 111
DB 1095 ddraeppsgkhndieemltisvptlvmevqpppdpiplevplntssnatcdhpetnavd 1154
QY 112 LGEALDTSSEPGQHPPAPETPIPGASWSISPTTGGLPWGTQTPTPYVEDKSGQGEDL 170
DB 1155 vpykl---pgldeevqpnllp-----rrpspyektrnrgl 1190
XX

```

```

RESULT 4
AAM47030
ID AAM47030 standard; Protein; 1206 AA.
XX
AC AAM47030;
XX
DT 06-JUL-1998 (first entry)
XX
DE Bovine N-proteinase.
XX
KW N-proteinase; cattle; collagen; antibody; rheumatoid arthritis;
fibrosis; Ehlers-Danlos disease; diagnosis; therapy.

```

```

XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 142 /note="encoded by CCC"
FT FT Misc-difference 207 /note="encoded by CCG"
FT FT Misc-difference 207 /note="encoded by CCG"
FT FT Misc-difference 464 /note="encoded by CCC"
FT FT Misc-difference 545 /note="encoded by CCC"
FT FT Misc-difference 549 /note="encoded by ATC"
FT FT Misc-difference 810 /note="encoded by CCG"
FT FT Misc-difference 810 /note="encoded by ATC"
FT FT Misc-difference 980 /note="no equivalent codon in nucleotide sequence"
XX
XX W09800555-A1.
XX
XX 08-JAN-1998.
XX
XX 03-JUL-1997; 97WO-US12427.
XX
XX 02-JUL-1997; 97US-0886333.
XX 03-JUL-1996; 96US-0021203.
XX
XX (COLL/) COLIGE A.
XX (LAPI/) LAPIERE C.
XX (PROC/) PROCKOP D J.
XX
XX Collige A, Lapiere C, Prockop DJ;
XX
XX WPI: 1998-086980/08.
XX N-PSDB; AAV06594.
XX
XX Polynucleotide sequence encoding human N-proteinase - used to
XX produce mature collagen in vitro and antibodies to treat fibrosis
XX and Rheumatoid arthritis
XX
XX Disclosure: Fig 5; 49pp; English.
XX
XX This polypeptide comprises human N-proteinase, the enzyme
XX responsible for cleaving N-propeptide from procollagen to produce
XX mature collagen. The amino acid sequence was deduced from a cDNA
XX sequence (see AAV06594) derived from bovine cDNA clones. Bovine
XX N-proteinases can be produced using a claimed method in which a
XX host cell is transformed or transfected with an N-proteinase
XX polynucleotide sequence, cultured in an appropriate culture medium,
XX and the N-proteinase is isolated from the medium. The N-proteinase
XX can be used for the production of mature collagen in vitro and for
XX the production of antibodies which may be used for diagnosis and
XX therapy of diseases including fibrosis and rheumatoid arthritis.
XX The N-proteinase may also be administered to treat a disease
XX resulting from insufficient production of N-proteinase, such as
XX Ehlers-Danlos disease. Human N-proteinases (see AAM47028-29) are
XX also disclosed.
XX
SQ Sequence 1206 AA;

```

```

Query Match 17.5%; Score 211.5; DB 19; Length 1206;
Best Local Similarity 29.1%; Pred. No. 6e-07;
Matches 52; Conservative 24; Mismatches 68; Indels 35; Gaps 4;
4 QWVPGSGPLHPINKISTEPCGDSVFCQMEVLDRCISIPGYHRLCCVSC----- 54
DB 1038 qwlstrpdpnspvqelskgrcgdksvfcrmevlsrscsipygnklckscnphnltlv 1097
QY 55 IKKASGPNPGDPGPTSLPFPSTPGSPPL---PGQDPADAAEPPEKPTGSGEDHOGHGRATQ 111
DB 1098 ddraeppsgkhndieemltisvptlvmevqpppdpiplevplntssnatcdhpetnavd 1157

```

```
QY 112 LPGAIDTSSPGTOHFAETPIPGASWSISPTTPGGLPWGWTOQTPHPVPEDKGQGGEDL 170
: | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1158 vpyk1----pg1edevqpnllp-----trpspyektrnqrldgel 1193
```

RESULT 5

ID AAB21254 standard; protein; 1201 AA.

DT	23-FEB-2001 (first entry)
XX	
DE	Human metalloproteinase KIAA0366.
XX	
KM	Human; KIAA0366; ADAMTS; metalloproteinase; ADAM;
KM	a disintegrin and metalloproteinase domain; thrombospondin domain;
KM	vaccine; nontropic; neuroprotective; antiparkinsonian;
KM	cerebroprotective; cytosolic; antiarthritic; immunosuppressive;
	Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
	autoimmune disease; brain tumour; brain injury.

OS Homo sapiens.
XX
XX
PN W0200053774-A2.
XX
XX
PD 14-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US06237.
XX
XX
PR 08-MAR-1999; 99US-0264585.
XX
PA (NEUR-) NEUROCRINE BIOSCIENCES INC.

PI Kelner GS, Clark M, Maki RA;
XX
DR WPT, 2000-594326/56.
DR N-PSDB; AAA95824.
XX
PT polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases
XX
PS Claim 13; Fig 8; 129pp. English.

The present sequence is human metalloproteinase KIAA0366. It shows homology to the ADAMTS family of proteins, which contain thrombospondin, integrin and metalloproteinase domains. ADAMTS polypeptides are useful for the manufacture of medicaments for treating conditions associated with neuroinflammation and/or neurodegenerative conditions associated with Alzheimer's disease and stroke. They are also useful for treating conditions associated with cell proliferative cell migration, inflammation and/or angiogenesis, such as cancer, arthritis and autoimmune diseases. They can be used to treat patients afflicted with an invasive tumour, a brain tumour or brain injury.

Query Match	Best Local Similarity	Matches	Score	DB	Length
16.9%;	38.7%;	12;	204;	21;	1201;
Conservative	12;	Mismatches	36;	Indels	44;
Gap					

QY	12	LHP	INK	IST	EPT	CD	RVS	IP	GY	HN	LC	VC	SC	IK	KA	SG	-----	6															
Db	1008	lppc	----	depcj	dk	sl	lcm	e	l	a	r	y	s	i	p	y	nk	lc	e	s	s	r	s	s	l	l	p	p	y	l	a	e	1

```
QY      61 -----PNEGPPG-----PTSLRPE-----SPLRGPDODPADAEPRGRKG I
          || - | ||||| : || | : | | | | | | |
Db     1064 thadvinspsdlprlvmptslvpyhsctpakmslsissvgynaya-atirpnskpdg L
QY     101 SEDHGGRATQLRGALDT-----SSPETQ 124
```

```
Db      1123 : | | | | :  
         anlrg--rsaqgasktvr|vtvpsspptk 1150
```

RESULT	6
AAB73550	
ID	AAB73550 standard; Protein; 1205 AA

DT 07-AUG-2001 (first entry)

DE	Human ADAM-type metalloprotease	MDT55, SEQ ID NO:10.
XX		
KW	Human: MDT55; ADAM-type metalloprotease; drug screening	
KW	A Disintegrin And Metalloprotease; cancer; arthritis.	

OS Homo sapiens

PN	JF2001017183-A
XX	
PD	23-JAN-2001.
MM	

PF	09-JUL-1999;	99JP-0196584.
XX		
PR	09-JUL-1999;	99JP-0196584.

PA (YAMA) YAMANOUCHI PHARM CO LTD
XX
DR WPI; 2001-275950/29.
DR N-PSDB; AAH20225.
DR

PT A new metal protease and its preparation for use as an anti-cancer and
 PI anti-arthritis therapeutic -
 XX
 PS Claim 1; Page 16-19; 22pp; Japanese.

The invention relates to the novel human ADAM A51 integrin And Metalloproteinase)-type metalloproteinases MDT54 (AAB7554d) and MDT55 (AAB73530). The metalloproteinases can be used for the treatment of cancers and arthritis. The invention also relates to the genes encoding MDT54 and MDT55, vectors and host cells containing the MDT54 or MDT55 genes, the recombinant production of MDT54 and MDT55, and antibody specific for MDT54 or MDT55, and methods of screening for compounds which modulate the activity of MDT54 and/or MDT55. The present sequence represents human MDT55.

sq	Sequence	1205 AA
----	----------	---------

Query Match	16.9%	Score 204	DB 22	Length 1205
Best Local Similarity	38.7%	Pred. No. 2e-06		
Matches	58	Mismatches	36	Indels 44; Gaps 8

12 LHPINKISSTEPGDRSVFCQMEVLDKICSIPIGHNALCCVSCINNNDS

```
Db      1012 1pbcm-----qebcjlguksllclqmevaylacs4pyjunaicccscam35bca4ppf1722acac-
Qy      61 -----PnGdPPG---PTSLPPF--STPG-----SPLRGPDADAAEAPGKPrNG 100
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Gb      1060 4ba3uivncvdprlvmntelunvbstrakkmjslssivvaanaav-affpnmskdd 1128
```

101 SEDHONGRATOT,GA,DT-----SSPGTO 124

```

      : | : | : | : | :
Db 1127 anlrq--rsagqgsktvr\vtvpsspptk 1154

```

ABG22373 standard: Protein: 1328 AA
ID

AC
XX
ABG22373;

DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22364.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS86560.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20: SEQ ID No 52732; 103pp: English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABC00010-ABC30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1328 AA:
SO

Query Match 15.6%; Score 188.5; DB 22; Length 1328;
Best Local Similarity 34.0%; Pred. No. 2.7e-05;
Matches 49; Conservative 12; Mismatches 70; Indels 13; Gaps 3;

QY 4 QWVPSGLPIPIKISTEPCGDRSVFQMEYLDRIKCIPIYHRLCVSC-----I 55
DB 1054 qwlsrpdpspriksskqhcqgdkslfcmeylsvcsigykrlldksclnynpntuv 1113
QY 56 KASGPNPGP---DPGPTSLPFFSPGSP-LPGQDPADAAEPGKPGSEDHQGRATQ 111
DB 1114 egrieppgkhdldfmpdlpvytvamevrpspsrplevlnaastnatedhpetnavd 1173
QY 112 LPGALDTSSSGTQHP-FAFETPTP 134
DB 1174 epykhngledvqpnllprrpsp 1197

RESULT 8
AAR80327
ID AAR80327 standard; Protein; 762 AA.
XX
AC AAR80327;
XX
DT 18-APR-1996 (first entry)
XX
DE Protein polymeric adhesion substrate 1-B.
XX
KW Pendent group; repeating unit; enzyme recognition site; sealant; fibrin;
KW enzymatic cross-linking; biocompatible material; structural integrity;
KW medical adhesive; wound closure; tissue repair; transglutaminase.
XX
OS Synthetic.
XX
PN W09523611-A1.
XX
PD 08-SEP-1995.
XX
PF 03-MAR-1995; 95WO-US02728.
XX
PR 03-MAR-1994; 94US-0205518.
XX
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI Cappelletto J;
XX
DR WPI: 1995-320413/41.
XX
PT Protein polymers comprising repeating units and sequences - capable
PT of enzyme-catalysed covalent bond formation useful as a
PT biocompatible material for wound closure and tissue repair
XX
XX Example 4; Page 45; 138pp: English.
XX
XX The amino acid sequence of the protein polymeric adhesion substrate
CC (PPAS) 1-B. The protein comprises 10 repeats of the PPAS1-B monomeric
CC repeat (AAR80326) which consists of the CLP 3.7 gene encoded sequence
CC (AAR80321) linked to the human fibrin cross-linking substrate peptide
CC 93.2 sequence (AAR80316). The polymers can be used in biological
CC systems where in situ formation of a biocompatible material with
CC structural integrity is required e.g. as medical adhesives and sealants
CC or for wound closure or tissue repair.
XX
XX Sequence 762 AA:
SO

Query Match 14.5%; Score 175; DB 16; Length 762;
Best Local Similarity 35.0%; Pred. No. 0.00014;
Matches 63; Conservative 8; Mismatches 51; Indels 58; Gaps 12;

QY 58 ASGP-----NPGPD-----PGPTSLPFFSPGSP-LPGQDPADAAEPGKPTG 100
DB 29 asdpmgagpfpqpgqlpsspgagpqtppqglp--gspgagpfpqpgqlpsspgagpqtppge 86
QY 101 SEDHQRATQLGALDTSPTGQHPFAPETP---IPGASWSISPTTGG--GLPW--GW 152
DB 87 gqgqhlhgagatq--agdvgspsg----apqtpgpgqlpsspg--apqtpgpgqlpsspg 136
QY 153 TQPTTP-----VPEDKGGGEGDLR-----PGTSLPADLPGRPPEPCHP 191
DB 137 psgpfpqpgqlpsspgagpqtppgeqgqhlhgagatqdvgsppagpqtppqglpsspgagpqt 196

RESULT 9
AAR80324
ID AAR80324 standard; Protein; 762 AA.
XX
AC AAR80324;
XX
DT 18-APR-1996 (first entry)

(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.

Cappello J, Ferrari FA;

WPT; 1998-387004/33.

Recombinant collagen-like polymers - useful for making gels, films, fibres, etc.

biological materials. The polymers have collagen-like properties, but may be easily expressed in micro-organisms in high efficiency. The new sequences can be tailored to give the desired properties.

Query Match	14.5%;	Score 175;	DB 19;	Length 762;
Best Local Similarity	35.0%;	Score 200;	DB 19;	Length 762;

Query Match	14.5%;	Score 175;	DB 19;	Length 762;
Best Local Similarity	35.0%;	Pred. No. 0.00014;		
Matches 63;	Conservative 8;	Mismatches 51;	Indels 58;	Gaps 12

```

QY 58 ASGP-----NCPD-----PGTSLPSTSTPGS-LPGQDPAADAAEPGKATG 100
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 29 ASDPAGAPGTPGPGGIPGSPAPGTPGPGGIP- GSPGAPGTCTPGGIPGSPGAPGTCTPGE 86
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 SEOHGGRATOLPGALDTPSSGPTGHPEAPETP---IPGASMSISTPTPG--GLPW--GW 150
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 87 GGGHHGGAKG---AGDVGSPG---APGTPGPGIPGSPG-APGTCTPGGIPGSPGA 136
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 153 TQTPTP-----VPEDKCGPGEDLRH-----PGTSLPADLGRPEPCHP 191
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 137 PGTPTPGGIPGSPGAPGTCTPGGEGGHHGGAKGAGDVGSPGAPGTCTPGGIPGSPGAPGTCTP 166
    || | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT	13	.
AAW57668		
ID	AAW57668	standard; peptide; 762 AA

AC AAW57668;

DT 27-AUG-1998 (first entry)

Collagen-like polymer.

KW Collagen-like polymer; synthetic polymer; fibre coating;
KW prosthetic device; catalytic substance.

PN US5773249-A.

XX 18-DEC-2000; 2000EP-0127668.
 PF
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (RYOW) KYOWA HAKKO KOSYO KK.
 XX
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH68079.
 DR

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Claim 17; SEQ ID NO: 6614; 246pp + Sequence Listing; English

The present invention provides a number of nucleotide and protein sequences from the *Corynebacterium* *Corynebacterium* glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of *corynebacterium* *corynebacterium*, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from *Corynebacterium* *corynebacterium*, and identifying a homologue of a gene derived from *corynebacterium* *corynebacterium*. *Corynebacterium* bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

Note: The sequence data for this invention is available from the European Patent Office under accession number EP 1 111 111.

CC needed to separate data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

Sequence 635 AA;

Query Match	13.2%;	Score 159.5;	DB 22;	Length 635;
Best Local Similarity	27.8%;	Pred. No. 0.0015;		
Matches 71;	Conservative 13;	Mismatches 78;	Indels 93;	Gaps 15

```

108 -----srlpagrnlptcvagsgsvapraasapavpnpaagvaapqsl 15
46 YHRLCVCSCIKKASPNNG---PDP---GPTSLPPFSPGSD---LPGPDPADAAP 94
56 pmdg-tspnknisthpapgnalpapgaglpitpakteqgavpptaaklpv-----pg 107
/ POSGPHLPINKTISTSEPTGDR-----SVEGMEVLYDRCSIP 45
UY -----

```

[illegible]

QY	180	---	DLGRPPPPCHP	191
Db	273	gapd	apgtpsi	paap 287

RESULT 16

ID	standard; Protein; 682 AA.
AA80330	

AC AAR80330;

DT 18-APR-1996 (first entry)

DE Protein polymeric adhesion substrate 1-C-

X	Pendent group; repeating unit; enzyme recognition site; sealant; fibrin.
KX	enzymatic cross-linking; biocompatible material; structural integrity;
KW	medical adhesive; wound closure; tissue repair; transglutaminase.
XX	Synthetic.
OS	
PM	
PM	
XY	WO9523611-A1.

PN	W095523611-A1
XY	

PD 08-SEP-1995
YY

PF 03-MAR-1995; 95WO-US02728.
VY

PR 03-MAR-1994; 94US-0205518.
XY

PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XY

PI Cappello J;
YY

DR WPI; 1995-320413/41.
XX

PT Protein polymers comprising repeating units and sequences - capable of enzyme-catalyzed conjugation to a bound format

PT biocompatible material for wound closure and tissue repair
XY

Example 4; Page 47; 138pp; English
xx

The amino acid sequence of the protein polymeric adhesion substrate (PPAS) 1-1-c, the protein comprises 10 repeats of the PNAS1-C monomeric repeat (AA8080329) which consists of the CIP 3.7 gene encoded sequence (AA8080321) linked to the human fibrin cross-linking substrate peptide 63-4 sequence (AA8080317). The polymers can be used in biological systems where *in situ* formation of a biocompatible material with structural integrity is required e.g. as medical adhesives and sealants or for wound closure or tissue repair.

SQ Sequence 682 AA;

Query Match	13.1%;	Score 158;	DB 16;	Length 682;
Best Local Similarity	37.6%;	Pred. No. 0.002;		
Matches 62; Conservative	9;	Mismatches 50;	Indels 44;	Gaps 14;

```

QY 58 ASGP-----NPGPD-----PQSTSLPEESTPGSF-LPGFDQDPADAAEPPGKPTG 100
    |||      |||      |||      |||      |||      |||      |||      |||
Db 29 ASDPMGAGPGLPQPGQLPQSPGAPGLPQPGLP--GSPGAPGLPQPGQLPQSPGAPGLPQPG 86

```

```
QY 101 SEDHQHGATQLPEALDITSSPTQ---HPFAETP----IPGASWSISPTTPG--GIPW 150
    :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 ak--qaqdvgs--pgapgt--pgpqqlpspgpqgtpgpqqlpspg--apgtlpgpqqlpg 139
```

QY 151 -GWTQTTPPVEE-DKGGPGEDLRHFGTSLPADDEGRPEPCAR 191
| | : | | | | | |
Db 140 spgagptlpgakagdgavspga----pctpgpqglpgspgapgtpr 180

RESULT 17

ID	standard; Protein; 682 AA
AAW49717	

AC AAW49717;

DT 12-OCT-1998 (first entry)

Protein polymer adhesive substrate PPA1-C.

Protein polymer adhesive substrate; PPSI-C; sealant;

[illegible]

Homo sapiens.

US5773577-A.

collagen-like polymer; synthetic polymer; fibre coating
prosthetic device; catalytic substance.

Job time: 5535 sec

Thu Aug 29 10:46:17 2002

us-09-938-330-22_copy_1040_1249.rag

Page 12

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 20:06:58 ; Search time 53.85 Seconds
(without alignments)
2234.055 Million cell updates/sec

Title: US-09-938-330-20

Perfect score: 6954

Sequence: 1 MAPRLALSYLLPLHCAALCT.....PTGPTLLCVLPDQSLRGHT 1252

Indexing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 71.1**
2: PIR1**
3: PIR2**
4: PIR3**
5: PIR4**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3326	47.8	1205	2	procollagen N-endo
2	1108.5	15.9	951	2	gene ADAMTS-1 prot
3	1072.5	15.4	2165	2	hypothetical prote
4	896.5	12.9	837	2	hypothetical prote
5	774.5	11.1	550	2	hypothetical prote
6	699.5	10.1	1558	2	hypothetical prote
7	699.5	10.1	1558	2	protein C37C3.6a I
8	677	9.7	1444	2	hypothetical prote
9	622	8.9	951	2	angiogenesis inhib
10	556.5	8.0	860	2	hypothetical prote
11	472.5	6.8	957	2	hypothetical prote
12	433	6.2	898	2	hypothetical prote
13	430	6.0	1059	2	hypothetical prote
14	420	5.7	826	2	hypothetical prote
15	398.5	5.2	571	2	monocyte surface a
16	361	5.2	609	2	cathepsin C precu
17	339.5	4.9	616	2	ecarin precursor
18	312	4.5	549	2	metalloproteinase
19	306.5	4.4	776	2	metalloproteinase
20	302	4.3	610	2	angiogenesis regulat
21	294.5	4.2	617	2	vascular apoptosis
22	294.5	4.2	814	2	metalloproteinase
23	294	4.2	1074	2	disintegrin-like m
24	276.5	4.0	1584	2	semaphorin F precu
25	273.5	3.9	478	2	brain-specific ang
26	272	3.9	903	2	fibronolytic metal
27	269.5	3.9	411	1	metlin alpha - mo
28	266.5	3.8	1042	1	fibronase (EC 3.4.
29	266	3.8	789	2	hypothetical prote
					androgen-regulated

Altting
C105
E10096

ALIGNMENTS

30	264	3.8	777	2	I48100
31	257.5	3.7	480	1	A30065
32	255	3.7	823	2	S18968
33	252.5	3.6	416	2	A37877
34	248	3.6	481	2	JC4342
35	245	3.5	429	2	A42972
36	240.5	3.5	905	2	S55059
37	240	3.5	407	2	S66260
38	240	3.5	478	2	JQ1301
39	238	3.4	1170	2	A40558
40	237.5	3.4	481	2	S43125
41	235.5	3.4	1170	1	TSHP1
42	234	3.4	660	2	S71949
43	234	3.4	734	2	JC4861
44	233.5	3.4	419	2	S41607
45	232	3.3	735	2	G02937

RESULT 1
118517
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine
N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18517
R:Collage, A.; Nusgens, B.V.; Lapierre, C.M.
submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A:Accession: T18517
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X9389; NID:e990769; PID:e228215; PIDN:CA65253.1
A:Experimental source: skin
A:Genetics:
A:Gene: PC I-NP
A:Function:
C:Description: catalyzes cleavage of the propeptides of type I and II collagens prior
C:Keywords: hydrolase; metalloproteinase

Query Match 47.8%; Score 3326; DB 2; Length 1205;

Best Local Similarity 52.1%; Pred. No. 1e-188;

Matches 641; Conservative 160; Mismatches 351; Indels 78; Gaps 20;

QY	7	LLSYLLPLHCAALCTAGSRTPELHLSGKLSIDYGVYPCSTDRGRFLSHVYSGPAASAG	66
DB	17	LLTLLPLPADARLAAADPPGPGHGAERTLAVPVTDAGRLVSHVSA-ATAPAG	75
QY	67	SNVVDPTPLPRHSHLRVARSPLHPCGTLMPCRGHSLYNNVVFCKELHLRPRR	126
DB	76	---VTRRRAPAPRPLS-GGSEEDPG-----RLFNIVVFRRDLHLRPNAR	121
QY	127	LVPVSSVEMQEDFRELPEOPLRQECVYTGVTGM-PAANAISNCDIAGLIRTDSTDF	185
DB	122	LVAAGTAVMOWGSCATRVPEPLGTCLVGDVAAGLAESSVALSNCDIAGLIMEEBEF	181
QY	186	FIEPLRGQOEKA-SGRHVVYRRRAVQOEAEEDG-----DLHNEAFGIGDLPNLTGL	239
DB	182	FIEPLRGKLAKEADGGRVHVYHPTTSR--PPLGGPQALDGTISDLSLSRALGV	239
QY	240	VGDLGDTERR-RRAKPGSVSIEVLVVDVSVAFFHKEHVQNVYLTIMNIVDEIYHDE	298
DB	240	LEERVNYSRRRRRAADDDVYIEVLGVDSVYGFHTEHVQKTLITIMNIVDEIYHDE	299
QY	299	SIGVHTNIVLVLLVNGVROSLSIERGNPSRSLEOVCRMAHSOORDPSHAHHHVF	358
DB	300	SIGAHINIVLVRIILLSYGKSKSLIETGNPSOSLENNCRMAVYLOOKPDPTDHDHHAIF	359

Qy	359	L7RQDDGPR-----GAPVVMGMBPLSCALNHNDEGSSAFVAHETGVYLGNEHDGGNGC	415
Db	360	L7RQDDGPRGSMOQAVLTGMCHAPVBSCTLNHNDEGSSAFVAHETGVYLGNEHDGGNGRC	419
Qy	416	ADBTSLGSVAAPLVQAAPAFHRFHMRSKOSKLETSRYLPSYDCLLDLPDPMPAPQPELPGIN	475
Db	420	GDEVRILGSIAPLVQAAPAFHRFHMRSKOSQOELSTRYLSHSDCLRDPPFTMDMPALPDLGLH	479
Qy	476	YSMDEOCRFDPGSGTYOTCLAFRTFEPBCQJLMSCHPDNRYFECKTKKGPRLDSTBCAPKWC	535
Db	480	YSMDEOCRFDPGSGTYOTCLAFRTFEPBCQJLMSCHPDNRYFECKTKKGPRLDSTBCAPKHC	539
Qy	536	FKGCHMKSPQEQYGOODGCMSSMTKFGSCRSRSGGVSRSRSGNPNPSPAYGPRPCGPM	595
Db	540	FKGCHMLTLPD-LIKRDKGMGAMSPRPGSGSRICGIVAFRTROCDNPPRPAHGRTCSGLA	598
Qy	596	FERYOVNCSEBCEPQTYEDFPAQOCCAKNSYTHONAKHSWPEPEPDDAOKELICQADT	655
Db	599	YDFOLCSODCPALADDFEEOCRQMDLTFEHDQAQHMHLPHENDAKERCHLCESEKET	658
Qy	656	GDAVEMNOVYHDTROCRSYDRPYSVCARGCVYGGDCEKSGMKADKCGVCGGSGNSHRT	715
Db	659	GEVYSKRMVHDTROCRSYDAPSLCYRGCRKRYGGDGYVIGSSQOEDKCGVCGGSGNSHCV	718
Qy	716	VKGTLGKASKQAGALKLVQIPAGARHIQTEALEKSPHRSVAKNOVTSFLLNPKGK-EAT	774
Db	719	VKGTFSSPKKLGIYIKKFEIPAGARHLLIOEADTSHNLAVNLETGKFLINNEEDVDPN	778
Qy	775	SRTTANGLEMEBAVEDAKSLKTSQPLREALAIIALPTEGGRSSSLATKYVLIHEDLP	834
Db	779	SKTILANGVEHREDEGRRTELDTMPRLHGTITLVIP--EDAKAISLTKYMIHEBSLN	836
Qy	835	LIGSNVNLLEBMDYTEWALKSWAPCSKACGGGIOFTKYCCRRRRDHMYORHLCDHKRP	894
Db	837	-VDNNVLEDDSVGYEWALKMSPCPCPGGGGSOFTKYCCRRRLDHKYVHROFCOSYKRP	895
Qy	895	KPIRRRCNOHPCSQPVAVTEEMGACRSRGKLGVOYRGIOCLPLPSNGHKKMPAKACAG	954
Db	896	KAIRTCPCQPCSQPVAVTEEMERPCSSCGKRGMOYRSKYVPLHNNTTRSVYHTRKOND	955
Qy	955	DREPARRPLCEVPBPADWIRLGAMSSOCSSATCGEGIOQROYVCRTNANSLGHCBGDRPDTQ	1014
Db	956	ARPERRARCNEELCPGMWRAGSMSCQSVTCNGTOBRPYLCRTADDSFGVCREEPERETAR	1015
Qy	1015	VCSLEPACGCHQNSTVADYAWELGTPEBGQWVPVPSGLHFIINKISSMCAABEPCOTGBRSVC	1074
Db	1016	ICRLPQPCRRNSTDSCSY-----VQMLSPDDNSPVQETSS---KGRQGGKSVFC	1065
Qy	1075	OMEVYLDRCYSIPGHRILCCVSC-----IKASGPNBGPDPGFTSLPPESTFGSPL-	1124
Db	1066	RMEVYLSRCSIPGINKLCSKSCNHDULPYDVARAPRPSGKHNDIELMPLTSLVPTLVME	1125
Qy	1125	--PGQVPADAAEPGPGRGSGEDHQHGRATQLPGLADTSSPGIOHFFAPETIPGASMSI	1182
Db	1126	VQPPGIGILEVPLMTSSNATNEDEHPETNAVDVYKI-----PGLEDEVQPNLIP-----	1175
Qy	1183	SPPTTGGJLPMWMTQTPPVYEDKQCPQPEDL	1212
Db	1176	-----RPSPYEKTRNQRIQEL	1192

RESULT 2
T00017
gene ADAMTS-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00017
C:Kuno, K.; Iizasa, H.; Ohno, S.; Matsushima, K.
Genomics 46, 466-471, 1997
A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene
A:Reference number: Z14055; MUID:98110583
A:Accession: T00017
A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA
A:Residues: 1-951 <RUN>
A:Cross-references: EMBL:AB001735; NID:g2809056; PID:BAA24501.1; PID:g2809057
A:Experimental source: strain 129SV
C:Genetics:
A:Gene: ADMTS-1
A:Introns: 226/1: 343/3; 388/1: 444/1; 539/3: 602/1; 660/3; 719/2
C:Superfamily: thrombospondin type 1 repeat homology
F:542-558/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match	15.9%;	Score 1108.5;	DB 2;	Length 951;
Best Local Similarity	29.2%;	Pred. No. 9,1e-58;		
Matches 300;	Conservative 141;	Mismatches 364;	Indels 223;	Gaps 41

Qy	87	RSLPHPGCLMPGAVGHSLYFVWVYGEKRLRLPBNRLVYPCSSVEMOEDRELEPQ	146
Db	53	RAPGHSDTT-----TRLRIDAFGQQLHKLPDSDGFLAPFTLQ-----TVGRS	96
Qy	147	PLRQ-----ECVYTGAVTGMGAANAISNCDLAGLIRDSDFTEP-----	189
Db	97	PGEAQNHLDPDGLAHCFYSGTAVGDPGSAALSLCEGVRGAFYLOGEGEFTIOPAPGVAT	156
Qy	190	--LERGOEKASQRT--HYVYRE-----AYQEWNAEPDGDJHNA-----	227
Db	157	ERLAPAVEEESRSARPOFHILRRRRSGGAKCGVADDETLPTSDRSRQNTROMVYR	216
Qy	228	-----FGLDLDLNLGLGVDDQLGDTERRRRRAKDGSIIEVLLVDDSVYRFHG	277
Db	217	DPYPDQACKPSGPGSI-----RRKRFVSSPRY-VETMLVADDSMADFHG-	259
Qy	278	EHVQNYVYTLTMNIYDELYHDESLGVHINIALYLVWVGRQSLSTERPNRSLEQYCR	337
Db	260	SGIKYLLTLTSVAREFKAPBSLINSTSLVAVYILVY-VEQKGEPEVTSNALTIRNQC	318
Qy	338	WAHSGORODPS--FAEHHDHVLELTRDFGFS-----GYAPYGMCHPLRSCALNHEDG	389
Db	319	W--QKQHNHSPSDRDEHYDTALTFTTRDLCGSHTCDTGLMADVGTCVCDSPRSYIEDG	376
Qy	390	ESSAFVIAHEFGHYLVGMHDQSGCAD---EHSLSVMAPLVQAAPHFPHMSRCSKLEL	446
Db	377	LOAAETTTAHELGHFVNMHD--DAKHCSLVNGVGDSHLMSMLSSLDHQPMSPCSAIYV	435
Qy	447	SKYLES--YDCLIDPFPFAMOPPELPIRGINYSMDECRDFGSGYOTCLAFRTPECKQ	504
Db	436	TSEFLDNGHECLMDKRPQNTI-KLPSDLPTLYADRNCQCTFEESKHCP--DAASTCTT	492
Qy	505	LMCSHPDNPYF-CKTKKGRPLDGTCECARKWKCFEGHCHIKMSPOITYGQ--DQGSWMTF	561
Db	493	LMCTTSGGLVLCQCKHFPMADPCTSCGBKWCYSGKVNKTDMKHRTAPVHGSMGPGFW	552
Qy	562	SGSCRSKGGVYRSRSCNNPSPAYGGRPCLGPFMEFYOVNSECP-GIYEDEFAOQCAK	620
Db	553	GHCSRLCGGGVOYTRKCDNVPYKNGKCYCGKRVYRNSCNIIDCCPDNNKCTTREOCSA	612
Qy	621	RNSYVYHONAKHS-----WVP-YEPDDAQCELIQOSADTGDVYFMNOVYHDCGTC	671
Db	613	HNEF-----SKASFPNEPTEVTEPKYAGVSPKDRCKLTCEAKGIGTFVYLOPKRVADGTPC	667
Qy	672	STRDPYSVARECEYVPGCCKEYSGMKADKCGVCGGDNSCHRTVGTGLKASKOQALX	731
Db	668	S-PDSTSVVQOQCVKACQDRLIDSKKFTDCGYVCGGNGSTCKKMSIY--TSTRGYHD	724
Qy	732	IVQIPAGAHIOIEALEKSPHRSVYKNOVTSF--ILNPKGKATSFTTAMGLEME--	786
Db	725	IYIIPAGANIEVK-----HRNQGSNNNSFLIARAADGYVILNNGFTLSYLEDOLTY	778
Qy	787	-----DAVEDAKESLTSGLPPLAIIALPTEGGPRSSLAKYVHHEDLLPLIGSN	839
Db	779	KGTVLKYGSSSALERISFSPLKEPRLTYVL-----WVG--	813
Qy	840	NVLLEMDTYEALXSMAPCSAACGGIOFTFKYCGRRRDHHNVORHLCDDKKRPFIR	899

Db 814 -----HALPKIKFTYF-----MKKTESF-- 833
 QY 900 RCNOHPCSPYVWTEWEGASRSGKLGVOYRGIOCLPLSLNGHAKWPAKACAGD-RPE 958
 Db 834 --NAIPTSE-WVIEWEGESKTCGS-GWOKRVQC--RDING--HPSACEAKEVKA 883
 QY 959 ARRCPLRVCPDQWRLGMSOCASATGEGIQOQOVYCRITANSLGHCEDRP-----DTY 1013
 Db 884 SFRPCADLPCE-HWQVGMSPSCSTCKGKRYKRLKCVSHDGVLSNESCDPLKKRKYI 942
 QY 1014 QVCSLPAC 1021
 Db 943 DECTLTOC 950

UT 3
 71
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 R:Accession: T21371; T24896
 R:Gadadity: S.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z19413
 A:Accession: T21371
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2165 <WIL>
 A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
 A:Experimental source: clone F25H8
 R:Gadadity: S.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z19949
 A:Accession: T24896
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2165 <WIL>
 A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
 A:Experimental source: clone T13H10
 C:Genetics:
 A:Gene: CESP:F25H8.3
 A:Map position: 4
 A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 84

Query Match 15.4%; Score 1072.5; DB 2; Length 2165;
 est Local Similarity 26.0%; Pred No. 3.2e-55;
 Matches 312; Conservative 176; Mismatches 425; Indels 287; Gaps 45;

Db 151 ECVYTGCVGMPGAAVAISNC--DGLAGLITDSTDFIEPLEKQ--QOEKASGRTH 204
 Db 168 DCYIRAHVKGVHOHST-VNLCSEDELGLGMLAPSGIHTVERLISNGTEHDCARRHQH 226
 QY 205 VVVRREAVOQEMAEPRDGLHNEAFGIGDRLNLLGLVGDLGT-ERK--RRNAKPGSYS 260
 Db 227 LVKFPDRMFKSD-----HLNSTVNETETTVATWQDMEDVIERKARSRAANSMDHY 281
 QY 261 IEVLLVDDSVVFNHCKEHVONYVLINMIVDEIHDESLGVINIALVRLIMVGRSL 320
 Db 282 VEVLVAVADTKMYEYHGRS-LEDYVLLFLSTVASITRHOSLRASIVVYKLLIVLKTENAG 340
 QY 321 SLIERKNPERSLEOVCRMAHSOQRODPSIAEHNDHIVFLTRQDFSPS-----GYAPVT 373
 Db 341 PRITLO-NAOQTLQDFCRKWOQYVNDPDDSSVQNHDAVALLTRKDLCRSOGKCDTLGLABELG 399
 QY 374 GCHPLRSCALNHEDFSFAFVIAHETGIVLGEHDDGONGCA-----D 417
 Db 400 TWCMDQKSCALIEDNLSAFLIAHGLGVFSPIDHDE-KKCSITMPVKNKYCKQSTKFD 458
 QY 418 EYSLGS---VMAFLVOAARHHRMSCKSLSTRYLPs---YDCLLDPPDPAMPQP- 468
 Db 459 KTOFQNNFHMAPLEYNTHPWSWSPSCAGMLERFLNNRQOTQCLFDQVERRYEDVF 518

QY 469 -PELPGINYSMEDECRDFEGSGYOTCLAFTEFPCOLWC-SHPDNPFCYTKKGRPLDG 526
 Db 519 VADDEPKRYADAHQCKEFVGPASELCPYMPT---CRRLMCATFEGSOMGCRTOHMPWADG 575
 QY 527 TECAQK--WCFKGIKMSPEQTYGDDGMSWTKSGSRSRGGGVYRARSCKNPPSP 584
 Db 576 TPDCDSRSMFCHHAGVLAPESLTKIDGQGMKRSWGESRFGCGGVGLDGLDSCSPAR 635
 QY 585 AYGGRCPLGPMFEYOVQVNSECPCTYEDFPAQOCARKNSTYVH-----QNAKHSWVP-YEP 639
 Db 636 RNGKGYCVGQREYRSCNTQCEPMDTOPYREVQSEFNNDIGLQVASTNTHTWVYKYN 695
 QY 640 DDDAQCELTQSDADTDGVVFMNOVHDTGRCYRDPYSVCARGECVPVGCDEKESGMA 699
 Db 696 VAPNERCKLYCRISGSAFYLIRDKYVDGTPCD-RNGDDICVAGACMPACDQHLSLTLR 754
 QY 700 DDKCGCGGDSHCRTYKATGLGKASKAGALAKIYOTPAGARHIOIAL-----EKSHRYV 755
 Db 755 RDKCGVCGGDDSSCKYVKGTFPNEG-TFGYNEWKIPAGSANIDIKQYNNKEDDNYL 813
 QY 756 VKNOVTSFILLPCKEATSR--TFPMGLEW-EDAVEDAKESLKTGPIPEAIALIA 810
 Db 814 SLRANGFELLNGHFQVSLARQOIAFDVLEYSQDAL---IERINGTPIRSDIYHV 870
 QY 811 LPTEGGPRSSLAIFY---YIHEDLLPLIG-----SNNVLE 844
 Db 871 L-SVGSHPPDISYEWYTAVPNAVIRPISALYLRVITDTWTECDRACGQOOSOKIMCL 928
 QY 845 EMDTY-----EMALKSNAPCSKACGGIGOTFKXGRR 877
 Db 929 DMSHHRSHDRNCONVLRPKQATRMKNDICSTWITTEDVSSCSAKCSGOKRQVSCVM 988
 QY 878 RDHNM--VQRLDCHKKRPPI----- 897
 Db 989 EGDROTPESEHLDRNSKPSDIASCYIDCSGRKMYNGEWTSCSETGSGNGKMHRSYCV 1048
 QY 988 -----RRCONHPCSPYVWTEWEGASRSGKLGVOYRGIOCLPL 938
 Db 1049 DSNRRVDESICGBOKEATREBCNRLPC--PRWYGHMSECSRSDCG-GVKRHAQCDIDA 1105
 QY 939 LSNQTH--KYMPAK-----AC-----RPE 958
 Db 1106 ADRETHSRSGPAQOTQENHCHACTWMOFGVMSDCSAKAGDVOYRDANCTDRHSVLP 1165
 QY 959 AR-----RCQLRVCPDQWRLGMSOCASATGEGIQOQOVYCRITAN-----SLGH 1004
 Db 1166 HRLCKMEKILTPCKHRESCP-KYKIGEMSQSVSCEDGSSRRVSCVSGANGTEYVMSLGC 1224
 QY 1005 CEGDRPDYOVCSL-----PACGNNQNTVADVWELGTDEGQVNPQSG 1049
 Db 1225 TADDRASHOTCMGLTGPCFWRNITMSACSVSGIGHRETTTC-TYRQSVASFCGDT- 1282
 QY 1050 PLHPINKISSMCAEPCTGDR-----SVFCOMEYLDRYGSIPIGYNHLCVSCIKKASG 1102
 Db 1283 ---KMPETSOTCHLLPCTSMKPSHMSPCSVTGSGLQGR-----SVSCTRSGEG 1328

RESULT 4
 T00355
 hypothetical protein KIA00688 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
 C:Accession: T00355
 R:ishikawa, K.; Nause, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998
 A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl
 A:Reference number: Z14142; MUID:98403880
 A:Accession: T00355
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-837 <ISH>
 A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190
 A:Experimental source: brain

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: C89114
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1558 <STO>
 A:Cross-references: GB:chr_V; PIDN:AMC25867.1; PID:g3294501; GSPDB:GN00023; CESP:C37C3.6
 C:Gene: C37C3.6a
 A:Gene: C37C3.6a
 A:Map position: 5

Query Match 10.1%; Score 699.5; DB 2; Length 1558;
 Best Local Similarity 30.0%; Pred. No. 2,4e-33;
 Matches 179; Conservative 86; Mismatches 225; Indels 107; Gaps 26;

```

551 ODGSSMTKFGSCRSRSGGVRSRSCNPNPSPAYGRCPLGPMFEYQVNSBEC-PGT 609
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    75 ETGNMGPPVPENECSRSGGVLKRCQSG-----DCTGASVRYISCNINACESGT 126

610 YEDFAOCCAKRNSYYHONAKHSWVPEPDDAOKCELICQSDADTGDVVF--MNOVYHD 667
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    127 --DFAEOCCAKFNDEALDGN-YHKWTPYK--GKNKCELVCK-PEGNGFYKKADKVV-D 178

668 GTRCSYRDPYVSCARGCVPGCDKEVGSMAKADKCGVCGGDSHCRITYKGLGAKSKQA 727
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    179 GTRCDSKS-NDICVDGECPLPGCDGLSSLKFKDCKGCKDGDSTCKTTEGRFDERNLSP 237

728 GALKLVQIPAGARHIQIEALEKSPHRSVKNQVYGS--FLINPKGKEATSRFTAMG--L 783
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    238 GYHDIKLPBEGATNRIKIQEARKSTNNIALKN---GSDHFLINNGLIQVEKEVEVGTTIF 294

784 EWEBAVEDAKESLKTSGPLPEALAILALPTEGGPRSSLAQYVHEDLLPLIGSNVLL 843
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    295 YVDDA---EPETLSAQGPLSEELTV-ALLFRKGSRTAIKYE-----STPL 337

844 EEMDTYEMALKSNAPCSKACGGGLOFTKGCRRRRDHMMYQRLCHDKKRPK-IRRCN 902
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    338 EEEVDYMKFDNMTPCYSVSGKGVQTRNLKIDGKKRGVDEDLCEENNATKPEFEKSC 397

903 QHPCSQPVWVTEWAGACSRSGKLGVOTRGIOCLPLSLNTHKVMYPAKACAGDREARRP 962
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    398 TVDC-EAEMFTGDMESCSSTCGDGGQOYRVYCHQVFANGRRTVYEDGCTYERPVVQOT 456

963 CLRVPCPAQMRILGMSOCSSTGCEGIIQOQVYCRTN-----ANSLGCEEDRPDT 1012
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    457 CNRPACP-EMQAGPWSACSEKCGDAFOYRSVYCRSEKEEGEKLAAADACPADDEKEDT 515

1013 VOVCSLPACGNGHONSTVRADVWL-----GTPEGQWVPQSGPLHPINKI----- 1057
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    516 ERTCNLGPCEG---LTFVYGEMLNLCRNDTEETREVTCKDSOGRAVPLEKCLVDNSTE 571

1058 -----SSMCAAP-----CTGD-----RSVFCQ-----MEVLDK-YC 1083
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    572 IPDTRSCATOPCEYEWTVSEWSKCTTECGHGKTRVYICAIHQNGGLEVVDESHC 628
  
```

RESULT 7

T34395

hypothetical protein C37C3.6b - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T34395; T34394
 R:Geisel, C.; Bradshaw, H.
 submitted to the EMBL Data Library, July 1996
 A:Description: The sequence of C. elegans cosmid C37C3.
 A:Reference number: Z21518
 A:Accession: T34395
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2167 <GEI>
 A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b
 A:Experimental source: strain Bristol N2; clone C37C3
 A:Accession: T34394

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1555, 'SKF' <GE2>
 A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a
 A:Experimental source: strain Bristol N2; clone C37C3
 C:Gene: C37C3.6b
 A:Gene: CESP:C37C3.6b; CESP:C37C3.6a
 A:Map position: 5
 A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 15

Query Match 10.1%; Score 699.5; DB 2; Length 2167;
 Best Local Similarity 30.0%; Pred. No. 3,6e-33;
 Matches 179; Conservative 86; Mismatches 225; Indels 107; Gaps 26;

```

551 ODGSSMTKFGSCRSRSGGVRSRSCNPNPSPAYGRCPLGPMFEYQVNSBEC-PGT 609
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    75 ETGNMGPPVPENECSRSGGVLKRCQSG-----DCTGASVRYISCNINACESGT 126

610 YEDFAOCCAKRNSYYHONAKHSWVPEPDDAOKCELICQSDADTGDVVF--MNOVYHD 667
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    127 --DFAEOCCAKFNDEALDGN-YHKWTPYK--GKNKCELVCK-PEGNGFYKKADKVV-D 178

668 GTRCSYRDPYVSCARGCVPGCDKEVGSMAKADKCGVCGGDSHCRITYKGLGAKSKQA 727
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    179 GTRCDSKS-NDICVDGECPLPGCDGLSSLKFKDCKGCKDGDSTCKTTEGRFDERNLSP 237

728 GALKLVQIPAGARHIQIEALEKSPHRSVKNQVYGS--FLINPKGKEATSRFTAMG--L 783
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    238 GYHDIKLPBEGATNRIKIQEARKSTNNIALKN---GSDHFLINNGLIQVEKEVEVGTTIF 294

784 EWEBAVEDAKESLKTSGPLPEALAILALPTEGGPRSSLAQYVHEDLLPLIGSNVLL 843
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    295 YVDDA---EPETLSAQGPLSEELTV-ALLFRKGSRTAIKYE-----STPL 337

844 EEMDTYEMALKSNAPCSKACGGGLOFTKGCRRRRDHMMYQRLCHDKKRPK-IRRCN 902
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    338 EEEVDYMKFDNMTPCYSVSGKGVQTRNLKIDGKKRGVDEDLCEENNATKPEFEKSC 397

903 QHPCSQPVWVTEWAGACSRSGKLGVOTRGIOCLPLSLNTHKVMYPAKACAGDREARRP 962
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    398 TVDC-EAEMFTGDMESCSSTCGDGGQOYRVYCHQVFANGRRTVYEDGCTYERPVVQOT 456

963 CLRVPCPAQMRILGMSOCSSTGCEGIIQOQVYCRTN-----ANSLGCEEDRPDT 1012
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    457 CNRPACP-EMQAGPWSACSEKCGDAFOYRSVYCRSEKEEGEKLAAADACPADDEKEDT 515

1013 VOVCSLPACGNGHONSTVRADVWL-----GTPEGQWVPQSGPLHPINKI----- 1057
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    516 ERTCNLGPCEG---LTFVYGEMLNLCRNDTEETREVTCKDSOGRAVPLEKCLVDNSTE 571

1058 -----SSMCAAP-----CTGD-----RSVFCQ-----MEVLDK-YC 1083
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    572 IPDTRSCATOPCEYEWTVSEWSKCTTECGHGKTRVYICAIHQNGGLEVVDESHC 628
  
```

RESULT 8

T18856

antigenesis inhibitor homolog - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T18856; T24653
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: Z19031
 A:Accession: T18856
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1444 <WIL>
 A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1
 A:Experimental source: clone C02B4
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1995

Query Match	8.0%	Score 556.5	DB 2	Length 860
Best local Similarity	24.9%	Pred. No. 3.4e-25		
Matches 227	Conservative 145	Mismatches 360	Indels 181	Gaps 52
QY	168	ISCCDGLLITDSTDFPIEFLNGCOEKASRTHTVYR--EAYQDEMA---EPDG	221	
Db	55	VSN-----GMVADADGGDHL-----TQSHQTH-VHHKRSLESDIECCPDANEDDY	102	
QY	222	DLHNEAFGLDPLNLLGLVGDOLDTERKKRHAKP--GSYSIEVLVVDSDV---VPF	274	
Db	103	PEDREIIMAKSM-LIRMYKD-----IIRAPRHRODITVELALVAFDDADWHDHFKKM	152	
QY	275	HGR---EHQAVLVLLIMLVDEIYNDESGLVNIATLVR---LIMYGROSLMIENGPP	328	
Db	153	YGAALAEENHHTFIMAVVNNIDVLTQRLDQRIKIKIVREILKNIPLHMAKKSNGV	212	
QY	329	SRLSEVQCMWASQRODRPSAEHNDHNVLTTRDF-----GRSGIATVYTGKCHPLNSC	382	

Search completed: August 28, 2002, 20:12:20
Job time: 322 sec

Thu Aug 29 10:46:11 2002

us-09-938-330-20.rpt

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 00:02:39 ; Search time 48.87 Seconds
(without alignments)
412,907 Million cell updates/sec

Title: US-09-938-330-22_COPY_1040_1249
Perfect score: 1206
Sequence: 1 PEGQWVPSGPHLPINKISS.....PTGTFICVLPDSQLRGHT 210
Indexing table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	211.5	17.5	1205	2 T18517	procollagen N-endo
2	180.5	15.0	839	2 T04859	extensin homolog F
3	160	13.3	319	2 T32250	hypothetical prote
4	156.5	13.0	309	2 T29293	hypothetical prote
5	155.5	12.9	317	2 T19143	hypothetical prote
6	154	12.8	577	2 T09024	proline-rich prote
7	151.5	12.6	210	2 B44984	collagen - nematod
8	151.5	12.6	286	2 S34665	collagen, cuticula
9	151	12.5	296	2 A31219	collagen, I - Caeno
10	150.5	12.5	316	2 T21928	hypothetical prote
11	149.5	12.4	301	2 T13134	hypothetical prote
12	148	12.4	325	2 T32248	hypothetical prote
13	149	12.3	1664	2 T18262	S-layer protein -
14	146	12.1	290	2 T24590	hypothetical prote
15	145.5	12.1	299	2 T20605	hypothetical prote
16	145.5	12.1	316	2 T19291	hypothetical prote
17	145.5	12.1	330	2 T05717	probable extensin
18	145.5	12.1	1255	2 T31065	diaphanous protein
19	145.5	12.1	1340	2 A39808	proteoglycan core
20	145.5	12.1	2327	2 T42630	aggreacan - bovine
21	145	12.0	297	2 T29525	hypothetical prote
22	145	12.0	299	2 T29596	hypothetical prote
23	145	12.0	392	1 T29596	hypothetical prote
24	145	12.0	392	1 T29596	hypothetical prote
25	144.5	12.0	313	2 T22828	collagen alpha 1(I)
26	144.5	12.0	356	2 T22827	hypothetical prote
27	144.5	12.0	888	2 S28791	hypothetical prote
28	142.5	11.8	360	2 T37285	collagen alpha 1(X
29	142.5	11.8	464	2 S22697	collagen dpy-2 - C
					extensin - Volvox

Description

close

30	142	11.8	295	2 A44984	collagen - nematod
31	141.5	11.7	299	2 T24833	hypothetical prote
32	141.5	11.7	316	2 S08169	collagen col-12 pr
33	141.5	11.7	316	2 S08170	collagen col-13 pr
34	140.5	11.7	266	2 T22706	hypothetical prote
35	140.5	11.7	299	2 T22705	hypothetical prote
36	140.5	11.7	1690	1 CGH01B	collagen alpha 4(I
37	140	11.6	403	2 S52796	collagen alpha 2(I
38	140	11.6	677	2 S23296	collagen alpha 2(I
39	139.5	11.6	435	2 T15143	hypothetical prote
40	139.5	11.6	515	2 F70904	hypothetical prote
41	139	11.5	298	2 UC1448	collagen col-34 -
42	139	11.5	337	2 T23794	hypothetical prote
43	138.5	11.5	228	2 A44982	collagen UCOL1 - p
44	138.5	11.5	302	2 A31921	collagen dpy-13 pr
45	138.5	11.5	314	2 T32985	hypothetical prote

ALIGNMENTS

RESULT 1
T18517
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine
N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T18517
R:Collge, A.; Nusgens, B.V.; Lapiere, C.M.
Submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A/Accession: T18517
A/Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1
A:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior
C:Keywords: hydrolase; metalloproteinase

Query Match 17.5%; Score 211.5; DB 2; Length 1205;
Best Local Similarity 29.1%; Pred. No. 2.2e-06;
Matches 52; Conservative 24; Mismatches 68; Indels 35; Gaps 4;

QY 4 OMWVPSGPHLPINKISSSTEPCGDSVFCQMEVLDRTCSIFGHRCCVSC----- 54
DB 1037 OMWVPSGPHLPINKISSSTEPCGDSVFCQMEVLDRTCSIFGHRCCVSC----- 54
QY 55 IKKASGPNPGRPPPTSLPFPSTPGSPPL---GPPDPADAEPFGKPTGSDHOGHGRATQ 111
DB 1097 DDBRPPSGKNDIDELMPTLSVPLVMEVQPPGICPLEVPLNNTSNNATDEHPETNAVD 116
QY 112 LFGALDTSPTGQHPFADETPPGCASWSISPTTGGCLPWGNTQTPPVPEBDKQPGEDL 170
DB 1157 VPKYI---PGLEDEVQPNLIP-----RRPSYEKTRNQRQDEL 1192

RESULT 2
T04859
extensin homolog F28A21.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C/Accession: T04859
R:Bevan, M.; Mueller, M.W.; Muenndieck, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15387
A/Accession: T04859
A:Molecule type: DNA

neutrons: 26/1; 117/1; 338/1; 411/3; 430/2; 498/2

61 PNPGRDPPTSLPPFSTPGSPLPGPDADAAEPGKPTGSEDHQHGRATQLPGALDTSS 120

QY 121 PGQHFPAPETPIPGASWSISPTTPGGLPWGWTQTPTPVPEDKGQPGED-LRHPTSLP 178

QY 179 ADLGRPRPECHPT 192

C:Species: Haemonchus contortus

R; Shamansky, L.M.; Pratt, D.; Boisvenue, R.J.; Cox, G.N.

Reference number: A44984; MUID:90136718

Accession: B44984
Status: preliminary

```
A;ResIdues: 1-210 <SHA>
A;Cross-references: GB:J04671; GB:J04670
```

51 CVSCIKKASGPNPGDDPGPTSLPPFESTBGSB-----I DGBODBDNNAEBBCEVTMGNN 100

CAFCPCGGRPGP-PGP-IGPPGEP--GTPGNPGAPGNDAPGGPPGPKGPPGPGK----- 105

Db 106 -----AGAPGA--AGQPGANAPSEPLVPGP PGPPTGPOGPPGPN GAP -GHPGAP-GA 155

198 I L F I U C F H P E F G N I D O R S A B I O T A M

: : | | | | | | | | | | | | | | | | | | | |

1000

RESULT 8

C; Superfamily: unassigned collagens

QY 7 PQGPIHPINKISSTEPCTGDRSVFCQMEVLDRYCISIPGYHRLCCVSCIKKASGPNPGPD 66

67 PGPTSLP-----PFS--TPGSLPGPD-----PADAAEPGK--PTGSEDHQ 100

QY 107 GRATOLPGALDTSS-----PGTQHPEAPETPI-----PGASWSISPTTPEGI.PWG 151

QY 152 WTQTPVPEDKQPGEDLRHPGTSLPADLGR-----PPECPHTG 193

collagen 1 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*

R;Kramer, J.M.; Cox, G.N.; Hirsh, D.
C, accession: A51219; 13/290; 13/291

A: Reference number: A90826; MUID:83050944

A;Residues: 1-296 <KRA>

A; Status: preliminary; translated from GB/EMBL/DBJ

A; Cross-references: EMBL:J01047; PIDN:AAA27988.1
Accession: M37201

A:Residues: 1-296 <KP3>
A: molecule type: DNA

A;Gene: col-1

unassigned collagenase

every match	12.56	score 151	DB 2	length 296
Best Local Similarity	36.88	Pred. No. 0.0047		
Matches 57; Conservative	3	Mismatches 59	Indels 36	Gaps 12

51 CVSCIKKASGPNPGR-----DPGPTSLP-----PFSTPGSLP-GPQDPADAAEPPGKPTG 100

Db 142 CKPCQGGPPGP-PGPPGAPGDGEGAGTGGRTDAAPSGPPGPPGAGAGAPG-PAG 199
Qy 101 SEDHGHGRATOLPGALDTSSTGTOHPFAPETP-IPGASWSISPTTGGLPWGWTOTPTPV 159
Db 200 E-----PG-----TPAISEPLTPGAFGEPSDGGPPGPPGAPG-GNDGPPGP- 241
Qy 160 PEDKGPGEEDLRHPTSLPADLPGRPPPCHPPTGT 194
Db 242 PGPKGAPGD-PGPG-----ADGSGSPGPPGPPGAGT 271

RESULT 10

T19288
hypothetical protein C15A11.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19288
R:Gardner, A.
Submitted to the EMBL Data Library, September 1996
Reference number: Z19103
Accession: T19288
Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-316 <WILL>
A:Cross-references: EMBL:Z79694; PIDN:CA80195.1; GSPDB:GN00019; CESP:T15A11.6
A:Experimental source: clone C15A11
C:Genetics:
A:Gene: CESP:C15A11.6
A:Map position: 1
A:Introns: 47/3
C:Superfamily: unassigned collagens

Query Match 12.5%; Score 150.5; DB 2; Length 316;
Best Local Similarity 30.9%; Pred. No. 0.0054;
Matches 56; Conservative 13; Mismatches 73; Indels 39; Gaps 12;

Qy 44 PGYH-RLCCVSCIKKASGPPNP-PDGPPTSLPSPSTPGSP-----PGQDPADAAP 94
Db 125 PSHQKQKCEIKCPGRLPGAPGAPGQGPKG--PNGNPGAPAGCGGQGPPEPPGASAGS 182
Qy 95 PGKPE-----TGSSEHGHGRATOLPGALDTSSTGTOHPFAPETP-IPGASWSISPTTGGT 148
Db 183 PGQAGAGNCGSPKSGQKRGRLPG--PSGAPSGGP--PGAGGQSGSNAAPGAPGPG- 237
Qy 149 PWGWTOTPTPVPEKDGQPGEDLRHPTSLPADLPGRPPPCHPPTGTTCVLRDSQLRG 208
Db 238 PAG-----PNGQPG-----HPGQD-----GQPGAPGNDGTGSDAAYCPCEPTRSSVLR 280
Qy 209 H 209
Db 281 H 281

RESULT 11

T21314
hypothetical protein F23H12.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T21314
R:Kershaw, J.
Submitted to the EMBL Data Library, June 1996
Reference number: Z19405
Accession: T21314
Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-301 <WILL>
A:Cross-references: EMBL:Z74472; PIDN:CAA98942.1; GSPDB:GN00023; CESP:F23H12.4
A:Experimental source: clone F23H12
C:Genetics:
A:Gene: CESP:F23H12.4
A:Map position: 5

A:Introns: 54/3; 285/2
C:Superfamily: unassigned collagens

Query Match 12.4%; Score 149.5; DB 2; Length 301;
Best Local Similarity 34.9%; Pred. No. 0.006;
Matches 59; Conservative 6; Mismatches 69; Indels 35; Gaps 13;

Qy 51 CVSCIKKASGPPNP-----DGPPTSLP-----PSPGSPLP--GQDPADAAPPKPTG 100
Db 147 CKPCQGGPPGP-PGPPGAPGDGEGAGTGGRTDAAPSGPPGPPGAGAGAPG-PAG 204
Qy 101 SEDHGHGRATOLPGALDTSSTGTOHPFAPETP-IPGASWSISPTTGGLPWGWTOTPTPV 159
Db 205 E-----PG-----TPAISEPLTPGAFGEPSDGGPPGPPGAPG-GNDGPPGP- 246
Qy 160 PEDKGPGEEDLRHPTSLPADLPGRPPPCHPPTGTTCVLRDSQLRG 208
Db 247 PGPKGAPGD-PGPGVGDGSGPPG-PGAPGTPGEGKIC--PKYCALDG 291

RESULT 12

T32248
hypothetical protein T15B7.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32248
R:Pailey, A.; Gattung, S.
Submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid T15B7.
A:Reference number: Z21139
Accession: T32248
Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-325 <PAU>
A:Cross-references: EMBL:AF022985; PIDN:AAB69961.1; GSPDB:GN00023; CESP:T15B7.4
A:Experimental source: strain Bristol N2; clone T15B7
C:Genetics:
A:Gene: CESP:T15B7.4
A:Map position: 5
A:Introns: 266/1
C:Superfamily: unassigned collagens

Query Match 12.4%; Score 149; DB 2; Length 325;
Best Local Similarity 28.3%; Pred. No. 0.0069;
Matches 52; Conservative 10; Mismatches 62; Indels 60; Gaps 9;

Qy 44 PGYHRLCCVSCIKKASGPPNP-PDGPPTSLPSPSTPGSP-----PGQDPADAAP 96
Db 128 PGF---CITCPAGAPG-PGPPGAPGPKGNNQGTGATAGSGRGPPGPPGAGAGSPG 182
Qy 97 KPTGSEHGHGRATOLPGALDTSSTGTOHPFAPETP-----IPGASWSISPTTGG 147
Db 183 QP-----GHPGSPGPNRGGRGRLPSPGPPGPPGAG 217
Qy 148 LPW-----GWTOTPTP--VPEDKGPGEEDLRHPTSLPADLPGRPPPCHPPTGTTCVLR 200
Db 218 PGQPGHSGAGSPGPGGPPGSPGPG-----HSGND--GVPGAGNPGSGGDAATPC 270
Qy 201 PRDS 204
Db 271 PARS 274

RESULT 13

T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993

A:Map position: 5

A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931
C:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <F0J>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 12.3%; Score 148; DB 2; Length 1664;
Best Local Similarity 26.8%; Pred. No. 0.035;
Matches 55; Conservative 14; Mismatches 78; Indels 58; Gaps 11;

1 PEGQWVPSGPHPI-NKISTECTGDRSVFCQMEVLDRYCSIPGHRICVSCIKKAS 59
802 PSDEPTSEPTPEPIPTDTPSDEPTPSDEPT-----PSD 835
QY 60 GPNFGPPGPTSLP-PSTPGSLP--GPQD---PADAAEPGKPTGSEDHQGRATQLP 113
DB 836 EPTPSDEPTSEPTPEPIPTDTPSDEPTPSDEPTPSDEPT-----P 886
QY 114 GALTSSPGQHPPAPETPIPGASWSISPT-----TPGILPWGWTQTPVPEDEKQGPGE 168
DB 887 TPSEPTPESE---TPPEPIPTDTPSDEPTPSDEPTPSDEPT-----TPSDEPTPSDEPTPE 939
QY 169 DLRHPTSLPADLPGRPPEPC-HPT 192
DB 940 TPEEP---IPTDTPSDEPTPSDEPT 961

RESULT 14
T24590
hypothetical protein T06E4.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T24590
R:Ilroy, C.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19910
A:Accession: T24590
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-290 <WIL>
A:Cross-references: EMBL:Z70756; PIDN:CAA94792.1; GSPDB:GN00023; CESP:T06E4.6
A:Experimental source: clone T06E4
A:Genetics:
A:Gene: CESP:T06E4.6
A:Map position: 5
A:Introns: 46/3
C:Superfamily: unassigned collagens

Query Match 12.1%; Score 146; DB 2; Length 290;
Best Local Similarity 32.0%; Pred. No. 0.0097;
Matches 56; Conservative 7; Mismatches 70; Indels 42; Gaps 11;

QY 51 CVSCIKKASGPNPDP-PGPTSLPSTPG-----SPLP-GPQDPADAEPKPTGSE 102
DB 128 CITCPAGAGPAGAPGAPGPG--PSGAPGQDAVGEPGAPGAPGAGAPGQ-AGAP 184
QY 103 DH-----QHCRATQLGALDTSSPGTQHFAFETPIPGASWSISPTTGGGLPWGWTQ 154
DB 185 GHGAPAGGQGSKRT--PG--PAGAPGQGP-AGGPGGPGGSGAGAPGAPGAP----- 234
QY 155 TPTPVPEDEKQGPGEDELRRHPTSLPADLPGRPPEPCHPPTGTTTLCVLPNDSOLRGH 209
DB 235 ---GAGCGGPGGQD-----GQCGAPGNDGAPGSDAAYCTCPARSQVYKH 276

RESULT 15
T20605
hypothetical protein F08G5.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T20605
R:Steward, C.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19299
A:Accession: T20605
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-299 <WIL>
A:Cross-references: EMBL:Z70682; PIDN:CAA94581.1; GSPDB:GN00022; CESP:F08G5.4
A:Experimental source: clone F08G5
C:Genetics:
A:Gene: CESP:F08G5.4
A:Map position: 4
A:Introns: 47/3
C:Superfamily: unassigned collagens

Query Match 12.1%; Score 145.5; DB 2; Length 299;
Best Local Similarity 35.4%; Pred. No. 0.011;
Matches 51; Conservative 7; Mismatches 51; Indels 35; Gaps 11;

QY 51 CVSCIKKASGPNPDP-----PGPTSLP-----PSTPGSLP-GPQD---PADAAEP-- 94
DB 136 CIKCPAGEAGP-AGPDGAPGAPGQGGQGPAGQGPAGPAGPAGGAGAPGAG 194
QY 95 PCKPTGSEDHQGRATQLGALDTSSPGTQHPFAPE-----TPIGASWSISPTTGG 146
DB 195 PGAP--GQDGQGR--TGILPGA-----PGAPGQGSNPGQDGAAGAPGAGPAGPAGPD 245
QY 147 GLPWGWTQTPVPE--DKGQGED 169
DB 246 GQPGAGQGGGAGPAGGAGGAGAD 269

RESULT 16
T19291
hypothetical protein C15A11.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19291
R:Gardner, A.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19103
A:Accession: T19291
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-316 <WIL>
A:Cross-references: EMBL:Z79694; PIDN:CAB01961.1; GSPDB:GN00019; CESP:C15A11.5
A:Experimental source: clone C15A11
A:Genetics:
A:Gene: CESP:C15A11.5
A:Map position: 1
A:Introns: 47/3
C:Superfamily: unassigned collagens

Query Match 12.1%; Score 145.5; DB 2; Length 316;
Best Local Similarity 30.9%; Pred. No. 0.011;
Matches 56; Conservative 12; Mismatches 74; Indels 39; Gaps 12;

QY 44 PGYH-RLLCVSCIKKASGPNPDP-DGPTSLPSTPGSLP-----PGQDPADAEP 94
DB 125 PSHHQKQKICPCGGLPFPAGVPQGPKG--PQNGAPAPGQGGGCGPAGPAGSAGS 182
QY 95 PGRP-----TGSEDHQGRATQLGALDTSSPGTQHPFAETP-IPGASWSISPTTGGGL 148
DB 183 PGQAGAPGPNPSPGKSGQGRGLPG--PSGAPGQGP--PGAPGQGSNAPGAPGAPG- 237
QY 149 PWGWTQTPVPEDEKQGPGEDELRRHPTSLPADLPGRPPEPCHPPTGTTTLCVLPNDSOLRG 208
DB 238 PAG-----PNGQPG--HPGQD---GQPGAPGNDGTGSDAAYCPCTRSSVLR 280

QY 209 H 209
Db 281 H 281

RESULT 17

T05717

probable extensin - barley (fragment)

C:Species: Hordeum vulgare (barley)

C>Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 20-Jun-2000

C:Accession: T05717

R:Doan, D.N.P.; Sturaro, M.; Olsen, O.A.

Submitted to the EMBL Data Library, July 1997

A:Description: Characterization of a nuclear cDNA encoding a probable extensin from de

A:Reference number: Z15429

A:Accession: T05717

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1330 <DOA>

A:Cross-references: EMBL:Z98204; PIDN:CAB10894.1

A:Title: The

A:Reference number: A34234; MUID:89380219

A:Accession: A34234

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 128-621 <ANT>

A:Cross-references: GB:J05028

R:Oldberg, A.; Antonsson, P.; Helnegard, D.

Biochem. J. 243, 255-259, 1987

A:Title: The partial amino acid sequence of bovine cartilage proteoglycan, deduced fro

A:Reference number: A27752; MUID:87270630

A:Accession: A27752

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 622-1340 <OLD>

R:Sandy, J.D.; Boynton, R.E.; Flannery, C.R.

J. Biol. Chem. 266, 8198-8205, 1991

A:Title: Analysis of the catabolism of aggrecan in cartilage explants by quantitation

A:Reference number: A39808; MUID:91217051

A:Accession: A39808

A>Status: preliminary

A:Molecule type: protein

A:Residues: 1-28;59-82;131-137, 'QSEF', 142-149;196-207;226-249;1137-1143;1252-1267;127

R:Perin, J.P.; Bonnel, F.; Jolles, P.

FEBS Lett. 206, 73-77, 1986

A:Title: Structural relationship between link proteins and proteoglycan monomers.

A:Reference number: A27751; MUID:87005253

A:Accession: A27751

A:Molecule type: protein

A:Residues: 29-58;74-130;174-175, 'A', 177-204;208-225 <PER>

R:Perin, J.P.; Bonnel, F.; Jolles, P.

FEBS Lett. 176, 37-42, 1984

A:Title: Sequence data concerning the protein core of the cartilage proteoglycan mono

A:Reference number: A91327; MUID:85027710

A:Accession: E29164

A:Molecule type: protein

A:Residues: 1230-1249 <PE2>

C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;

C:Keywords: glycoprotein

F:8-28/Domain: link protein repeat homology (fragment) <LNK1>

F:29-58/Domain: link protein repeat homology (fragment) <LNK2>

F:80-146/Domain: link protein repeat homology (fragment) <LNK3>

F:167-248/Domain: link protein repeat homology (fragment) <LNK4>

F:1130-1250/Domain: C-type lectin homology <LCH>

F:1257-1313/Domain: complement factor H repeat homology <FHD>

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A:Molecule type: protein

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A:Molecule type: protein

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A:Molecule type: protein

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

QY 111 QLPALDTSSPQHPAPERTIPGASMSISPTTGGILPMGWOTQTPRPVEDKGOPEDEL 170
Db 677 -LPG-----GTGIP-PPPPPLPG-----SVGVPPPLPG-GGGGLPPPPPPPGAGGIP 723

QY 171 RHPGSLPADLP---GRPPPCHPPTGFTLLCVLRDSQLR 207
Db 724 PPGMGVPPPPPPGFGVPAFVLPFGILPKKVKYKREVOLR 763

RESULT 19

A39808

proteoglycan core protein, cartilage - bovine (fragments)

N:Alternate names: aggrecan; aggregating cartilage proteoglycan

C:Species: Bos primigenius taurus (cattle)

C>Date: 20-Mar-1992 #sequence_revision 23-Mar-1995 #text_change 13-Aug-1999

C:Accession: A34234; A27752; A39808; A27751; E29164; B27751; C27751; D27751; E27751;

R:Antonsson, P.; Helnegard, D.; Oldberg, A.

J. Biol. Chem. 264, 16170-16173, 1989

A:Title: The keratan sulfate-enriched region of bovine cartilage proteoglycan consist

A:Reference number: A34234; MUID:89380219

A:Accession: A34234

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 128-621 <ANT>

A:Cross-references: GB:J05028

R:Oldberg, A.; Antonsson, P.; Helnegard, D.

Biochem. J. 243, 255-259, 1987

A:Title: The partial amino acid sequence of bovine cartilage proteoglycan, deduced fro

A:Reference number: A27752; MUID:87270630

A:Accession: A27752

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 622-1340 <OLD>

R:Sandy, J.D.; Boynton, R.E.; Flannery, C.R.

J. Biol. Chem. 266, 8198-8205, 1991

A:Title: Analysis of the catabolism of aggrecan in cartilage explants by quantitation

A:Reference number: A39808; MUID:91217051

A:Accession: A39808

A>Status: preliminary

A:Molecule type: protein

A:Residues: 1-28;59-82;131-137, 'QSEF', 142-149;196-207;226-249;1137-1143;1252-1267;127

R:Perin, J.P.; Bonnel, F.; Jolles, P.

FEBS Lett. 206, 73-77, 1986

A:Title: Structural relationship between link proteins and proteoglycan monomers.

A:Reference number: A27751; MUID:87005253

A:Accession: A27751

A:Molecule type: protein

A:Residues: 29-58;74-130;174-175, 'A', 177-204;208-225 <PER>

R:Perin, J.P.; Bonnel, F.; Jolles, P.

FEBS Lett. 176, 37-42, 1984

A:Title: Sequence data concerning the protein core of the cartilage proteoglycan mono

A:Reference number: A91327; MUID:85027710

A:Accession: E29164

A:Molecule type: protein

A:Residues: 1230-1249 <PE2>

C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;

C:Keywords: glycoprotein

F:8-28/Domain: link protein repeat homology (fragment) <LNK1>

F:29-58/Domain: link protein repeat homology (fragment) <LNK2>

F:80-146/Domain: link protein repeat homology (fragment) <LNK3>

F:167-248/Domain: link protein repeat homology (fragment) <LNK4>

F:1130-1250/Domain: C-type lectin homology <LCH>

F:1257-1313/Domain: complement factor H repeat homology <FHD>

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A:Molecule type: protein

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A:Molecule type: protein

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

A:Molecule type: protein

A:Residues: 1-1255 <WAT>

A:Cross-references: EMBL:J096963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A:Note: binds to GTP-bound form of Rho and binds to profilin

A:Accession: T31065

QY 5 WVPQSGPLPINKISSTEPCTGDRSVFCOMEVLDRC-----SIFGYHRLCCVSCIKK 57
Db 200 WLADGSLRIRPI--VTPRACGGDKRGVTRIVLYLPTGTGLD--PLSRHACFGVGSAA 254

```

QY 56 ---KKASGPRPGD-----PG-----PPLSPPFS-----TP 78
      :: | | | |
Db 255 PSEEEEGSAPRAGDAVEEMWYTVGSGVAAYPIGGETTAIRGFVPEPKTEWELAYTP 314
      : : : :
QY 79 GSPLEPQOPADAAEPKPRG--SEDHQG-RATGLPADLTSSPQHFAETPIPG 135
      : : : :
Db 315 AGTLPLPGIP-----PTMPPTGATEHTHTGSPATVPSASEKPRP-SEEPFPDEEPP- 367
      : : : :
QY 136 ASWSISPTTPGGLPWGWTQTPTPVEBKSGQSGDELRHPTSLPADLRPREPCHP 192
      : : : :
Db 368 ---SEKPPPEEL-----PPSKPPPSKPPPSSE-----PPSEKPP-PPPEELFPS 410

```

RESULT 20
630

Species: bos - bovine
 Species: Bos primigenius taurus (cattle)
 Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
 C:Accession: T42630
 R:Herling, T.M.; Kollar, J.; Huynh, T.D.
 Submitted to the EMBL Data Library, September 1996
 A:Description: Complete coding sequence of bovine aggrecan: comparative structural analysis
 A:Reference number: Z22182
 A:Accession: T42630
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2327 <HCR>
 A:Cross-references: EMBL:U076615; NID:g1730259; PID:g1730260; PIDN:AAB38524.1
 A:Experimental source: articular chondrocytes
 A:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EC
 C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein

Query Match	12.18;	Score 145.5;	DB 2;	Length 2327;
Best Local Similarity	37.09;	Score 145.5;	DB 2;	Length 2327;

Local Similarity 27.06; Pred. NO. 0.069;
Matches 64; Conservative 27; Mismatches 71; Indels 75; Gaps 17.

matchines 84; conservative 27; Mismatches 71; Indels 75; Gaps 17;

QY 5 WPOSGLHPINKJSTSEPTG---RSYFC--QMEVLDRCYSIPGYHRLCC--VSCI 55
|:::|||||::|

Db 635 WLADGSLRYP I - - VTPRPACGGDKPGVRTVYLLYPNQ TGLLD - - PLSRHHAFCFRGVSAA 689

QY 56 ---KKASGPNCPD-----PC-----PTSLPPS-----TP 78
::|||

Db 690 P S P E E E G S A P T A G P D V E E M V T Q V G P G V A A V P I G E T T A I P G F T V E P E N K T E M E L A Y T P 749

79 GSPLPPODPADAAEPBGKPTG--SEDHONG-RATOI PGAT DTSEBCTOUNE,DEEETDQ 122

[illegible]

130 AGILLPLGIP-----PTWPTGEATEHTEGPSATEVPSASEKPP-SEEPFPEEPF- 802

QY 136 ASWSISPTTPGCLPWGWTQTPTPVPEDEKQPGEDLRHPTSLPADLPGRPPPECHPT 192

```

Db      803  ---SEKFPPEEL-----FPSEKFPPESEKFPPEF-----DPSFKV-EDDEETEND 846
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
          : : : : : : : : : : : : : : : : : : : : : : : : : :
          |::| | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: August 29, 2002, 00:09:27
Job time: 408 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 00:06:29 ; Search time 37.54 Seconds
(without alignments)
216.598 Million cell updates/sec

Title: US-09-938-330-22_COPY_1040_1249
Perfect score: 1206
Sequence: 1 PEGQWVPSGPIHPINKISS.....PTGTFITLCVLPKDSQLNGHT 210

ing table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	211.5	17.5	1205 1	AT2S2_BOVIN
2	207.5	17.2	1211 1	AT2S2_HUMAN
3	204	16.9	1201 1	AT2S3_HUMAN
4	155.5	12.9	317 1	Y035_CAEEL
5	151.5	12.6	210 1	CAC2_HAEEO
6	151	12.5	296 1	CC01_CAEEL
7	148.5	12.3	1664 1	GPL_CLOTH
8	148	12.3	1664 1	GPL_CLOTH
9	145.5	12.1	1255 1	DIA1_MOUSE
10	145.5	12.1	1255 1	DIA1_MOUSE
11	145	12.0	331 1	PRP1_HUMAN
12	145	12.0	1466 1	PRP1_HUMAN
13	142	11.8	295 1	CAC3_HAEEO
14	141.5	11.7	316 1	CC12_CAEEL
15	141.5	11.7	316 1	CC13_CAEEL
16	140.5	11.7	266 1	YXWK_CAEEL
17	140.5	11.7	1690 1	CC44_HUMAN
18	139	11.5	298 1	CC34_CAEEL
19	138.5	11.5	302 1	CCDC_CAEEL
20	138.5	11.5	1183 1	DRPL_RAT
21	138.5	11.5	1464 1	CA13_MOUSE
22	138.5	11.5	1496 1	CA23_HUMAN
23	137.5	11.4	503 1	WAIP_HUMAN
24	137	11.4	1248 1	CA1A_HUMAN
25	136	11.3	251 1	DIA1_HUMAN
26	136	11.3	1459 1	PRP2_MOUSE
27	135.5	11.2	754 1	CA54_CANA
28	135.5	11.2	1097 1	S24C_ARATH
29	135	11.2	301 1	CC02_CAEEL
30	134.5	11.2	279 1	Y091_NPYOP
31	134.5	11.2	1464 1	CA11_HUMAN
32	134	11.1	2944 1	CA17_HUMAN
33	134	11.1	2944 1	CA17_HUMAN

34	133.5	11.1	1069 1	S24B_ARATH
35	133.5	11.1	1262 1	CA13_CHICK
36	133.5	11.1	1707 1	CA2A_MOUSE
37	133	11.0	449 1	APG_BRANA
38	132.5	11.0	360 1	CCD2_CAEEL
39	132.5	11.0	779 1	CA11_BOVIN
40	132.5	11.0	1804 1	CA1B_MOUSE
41	132	10.9	3063 1	CA1C_HUMAN
42	131.5	10.9	707 1	SEPO_HUMAN
43	131.5	10.9	1453 1	CA11_MOUSE
44	131	10.9	278 1	YLS6_CAEEL
45	131	10.9	1669 1	CA14_MOUSE

ALIGNMENTS

RESULT 1	ID	AT2S2_BOVIN	STANDARD:	PRT: 1205 AA.
AC	P79331	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)			
DE	(Procollagen I/II amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI).			
GN	ADAMTS2 OR NPI.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovine; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skin;			
RC	MEDLINE=9725960; PubMed=9122202;			
RA	Colige A., Li S.W., Steron A.L., Nusgens B.V., Prockop D.J.,			
RA	Lapierre C.M.;			
RT	"CDNA cloning and expression of bovine procollagen I N-proteinase: a new member of the superfamily of zinc-metalloproteinases with binding sites for cells and other matrix components."			
RT	Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).			
RN	[2]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=95348096; PubMed=7622483;			
RA	Colige A., Beschlin A., Samyn B., Goebels Y., Van Beeumen J.,			
RA	Nusgens B.V., Lapierre C.M.;			
RT	"Characterization and partial amino acid sequencing of a 107-kDa procollagen I N-proteinase purified by affinity chromatography on immobilized type XIV collagen."			
RT	J. Biol. Chem. 270:16724-16730(1995).			
CC	- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN COLLAGEN BIOSYNTHESIS.			
CC	- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(I) and alpha-2(I) chains at Ala-1-Gln.			
CC	- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV.			
CC	- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).			
CC	- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRA LEVELS WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.			
CC	- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-I DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.			
CC	- PTM: THE N-TERMINUS IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY			
CC	- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY			

DR EMBL: AJ003125; CAA05880.1; -
 DR MEROPS: M12.301; -
 DR MIM: 604539; -
 DR MIM: 225410; -
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR002870; Pep_M12B-propep.
 DR InterPro: IPR001590; Reptolysin.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR00130; Zn_MTPeptidase.
 DR Pfam: PF01562; Pep_M12B-propep; 1.
 DR Pfam: PF01421; Reptolysin; 1.
 DR Pfam: PF00090; TSP1; 4.
 DR SMART: SM00209; TSP1; 4.
 DR PROSITE: PS50215; ADAM_MEPRO; 1.
 DR PROSITE: PS50092; TSP1; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 DR PROSITE: PS00427; DISINTEGRIN; FALSE_NEG.
 DR Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
 KM Alternating splicing.
 FT SIGNAL 1 29 POTENTIAL.
 FT PROPEP 30 253 BY SIMILARITY.
 FT CHAIN 254 1211 ADAMTS-2.
 FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 409 409 BY SIMILARITY.
 FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 480 560 DISINTEGRIN-LIKE.
 FT DOMAIN 561 617 TSP TYPE-1 1.
 FT DOMAIN 618 722 CYS-RICH.
 FT SITE 723 851 CELL ATTACHMENT SITE (POTENTIAL).
 FT DOMAIN 851 851 SPACER.
 FT DOMAIN 852 911 TSP TYPE-1 2.
 FT DOMAIN 912 974 TSP TYPE-1 3.
 FT DOMAIN 975 1030 TSP TYPE-1 4.
 FT DOMAIN 1031 1031 POLY-ALA.
 FT DOMAIN 185 188 POLY-GLU.
 FT CARBOHYD 112 112 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 251 251 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 993 993 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 1031 1031 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 1098 1098 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 1145 1145 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 1150 1150 N-LINKED (GLCNAC) (POTENTIAL).
 FT VARSPLIC 544 566 HCFGHCITLTPDLKRDGSGMA -> FRGGAHAHCYPS
 TLAGGGRWIA (IN ISOFORM SPNFI).
 VARSPLIC 567 1211 MISSING (IN ISOFORM SPNFI).
 SEQUENCE 1211 AA; 134722 MW; BECEEF25C23CAD2D CRC64;

Query Match 17.2%; Score 207.5; DB 1; Length 1211;
 Best Local Similarity 34.7%; Pred. No. 2, 1e-05;
 Matches 50: Conservative 14; Mismatches 67; Indels 13; Gaps 3;

QY 4 QWVPSGPHLPNKSSTEPTGDSVFCQMEVLDRCSIPGYHRLCVCSC-----I 55
 DB 1043 QWLSRPDPSPIRKISGHCQGDKSIKRMELVSRCSIIPIGYKLSCKSNLYNNLTNV 1102
 QY 56 KASGPNPGP---DPGFTSLPPSTPGSLPGQPADAAAEFGKRTGSEDHQHGATQ 111
 DB 1103 EKRIPPEPGKHNDIVFMTLPVPEVAMEVRSPSTPLEVPLNASTNATEDHETNAVD 1162
 QY 112 LFGALDTSSPGTQHP-FAPETPIP 134
 DB 1163 EBYKIHGLEDEVQPNLIPRRSP 1186

RESULT 3
 ATSS3_HUMAN STANDARD; PRT; 1201 AA.
 ID ATSS3_HUMAN
 AC 015072;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (Fragment).
 GN ADAMTS3 OR KIA0366.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.;
 RL DNA Res. 4:141-150(1997).
 CC -1- CORRECTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 CC MATRIX (BY SIMILARITY).
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FORIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB002364; BAA20821.1; -
 DR MEROPS: M12.220; -
 DR MIM: 605011; -
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR002870; Pep_M12B-propep.
 DR InterPro: IPR001590; Reptolysin.
 DR InterPro: IPR000884; TSP1.
 DR Pfam: PF01562; Pep_M12B-propep; 1.
 DR Pfam: PF01421; Reptolysin; 1.
 DR Pfam: PF00090; TSP1; 4.
 DR SMART: SM00209; TSP1; 4.
 DR PROSITE: PS50215; ADAM_MEPRO; 1.
 DR PROSITE: PS50092; TSP1; 2.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 DR PROSITE: PS00427; DISINTEGRIN; FALSE_NEG.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT NON_TER 1 1
 FT SIGNAL 1 245 POTENTIAL.
 FT PROPEP 246 1201 BY SIMILARITY.
 FT CHAIN 246 1201 ADAMTS-3.
 FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 395 395 BY SIMILARITY.
 FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 404 404 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 466 546 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 547 603 DISINTEGRIN-LIKE.
 FT DOMAIN 604 708 TSP TYPE-1 1.
 FT DOMAIN 709 840 CYS-RICH.
 FT DOMAIN 841 898 SPACER.
 FT DOMAIN 899 961 TSP TYPE-1 2.
 FT DOMAIN 962 1013 TSP TYPE-1 3.
 FT DOMAIN 1013 1245 TSP TYPE-1 4.
 FT DOMAIN 1245 1245 POLY-ARG.

QY 104 HOGRATOLPGALDTSSTPGTHPFAETP-----IPGASMSISPTTPGGLPMGWTOTPTPV 159
 Db 106 -----AGAGCA--AGCPGANAPSEPLVECPGPPGPTGPEGPPGNGAP-GHPGAP-CA 155
 QY 160 PEDKGPGEEDLRHPTGSLPADLGRPPPCHPPTGTTLC 198
 Db 156 PGEKGPGRGD-GHPGAPGNAGHPGPGGP-GRPPGKGVG 192

RESULT 6

CC01_CAEEL STANDARD; PRT: 296 AA.
 ID CC01_CAEEL P08124;

01-AUG-1988 (Rel. 08, Created)
 01-AUG-1988 (Rel. 08, Last sequence update)
 01-FEB-1996 (Rel. 33, Last annotation update)
 Cuticle collagen 1.
 SQT-3 OR COL-1 OR DPY-15.
 GN Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=85105075; PubMed=2578467;
 RA Kramer J.M., Cox G.N., Hirsch D.;
 RT "Expression of the Caenorhabditis elegans collagen genes col-1 and
 col-2 is developmentally regulated."
 RL J. Biol. Chem. 260:1945-1951(1985).
 RN (2)

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=83050944; PubMed=7139711;
 RA Kramer J.M., Cox G.N., Hirsch D.;
 RT "Comparisons of the complete sequences of two collagen genes from
 Caenorhabditis elegans."
 RL Cell 30:599-606(1982).

CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 LINKS.

CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
 COLLAGENS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL: J01047; AAA27988.1; -;
 DR EMBL: V00147; CAA23463.1; -;
 DR PIR: A31219; A31219.
 DR InterPro: IPR002486; COL_cuticle_N.
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF01484; COL_cuticle_N; 1.
 DR Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
 KW DOMAIN 100 129 TRIPLE-HELICAL REGION.
 FT DOMAIN 148 174 TRIPLE-HELICAL REGION.
 FT DOMAIN 178 204 TRIPLE-HELICAL REGION.
 FT DOMAIN 213 278 TRIPLE-HELICAL REGION.
 SO SEQUENCE 296 AA; 28574 MW; 843B169D2C44526A CRC64;

Query Match 12.5%; Score 151; DB 1; Length 296;
 Best Local Similarity 36.8%; Pred. No. 0.011;

Matches 57; Conservative 3; Mismatches 59; Indels 36; Gaps 12;

QY 51 CVSCIKKASCPNPGP-----DPGTSLSL-----PFSPTGSLP-GPODDADAEPPGKPTG 100
 Db 142 CKPCPGPPPP-PGPGAPGDPDEAGTPGRPGTDAAPSGSPGRGPPGAGEGAPG-PAG 199
 QY 101 SEDHOGHGRATOLPGALDTSSTPGTHPFAETP-IPGASMSISPTTPGGLPMGWTOTPTPV 159
 Db 200 E-----PG-----TPAISEPLTPGAPGEPPDSCGPPGPBPAP-GNDGPPGP- 241
 QY 160 PEDKGPGEEDLRHPTGSLPADLGRPPPCHPPTG 194
 Db 242 PGPAGGAPD-GPPG-----ADGSGPPGPPAGT 271

RESULT 7

GPI_CHURE STANDARD; PRT: 555 AA.
 ID GPI_CHURE AC 09FP06; 003927;

DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
 glycoprotein 1).
 GN GPI.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadales; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN (1)

RP SEQUENCE FROM N.A.
 RX PubMed=11258910;
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
 RA Goodenough U.W.;
 RT "Glycosylated polypurine II rods with kinks as a structural motif in
 plant hydroxyproline-rich glycoproteins."
 RL Biochemistry 40:2978-2987(2001).

RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=91017504; PubMed=1699225;
 RA Adair W.S., Apt K.E.;

RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
 encoding cell wall hydroxyproline-rich glycoproteins."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).

CC -1- FUNCTION: Major component of the outer cell wall W6 (crystalline)
 layer.

CC -1- SUBUNIT: Associates with GP2 and GP3.

CC -1- PTM: N-glycosylated and O-glycosylated.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL: AF309494; AAG45420.1; -;
 DR EMBL: M58496; AAG459706.1; ALT-SEQ.
 DR GlycoSuiteDB: 09FP06; -;
 DR InterPro: IPR003882; Pistill_extensin.
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENSIN.
 DR PRINTS: PR01218; PSTLXTENSIN.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
 FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSX REPEATS.
 FT DOMAIN 259 279 POLY-PRO.
 FT CARBOHYD 399 399
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;


```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U96963; AAC53280.1; -.
DR      MCD; MGI:1194490; Diapl.
DR      InterPro; IPR003104; FH2.
DR      Pfam; PF02181; FH2; 1.
DR      SMART; SM00498; FH2; 1.
DR      K01 Colled coil; Repeat.
DR      DOMAIN 460 562
DR      DOMAIN 63 260
DR      DOMAIN 157 457
DR      DOMAIN 586 747
DR      DOMAIN 752 1197
DR      DOMAIN 1027 1179
DR      DOMAIN 1180 1194
DR      DOMAIN 1196 1199
DR      SEQUENCE 1255 AA; 139343 MW; 09404164873CA7C1 CRC64;
      COILED COIL (POTENTIAL).
      GBD.
      FH3.
      FHL (PRO-RICH).
      FH2.
      COILED COIL (POTENTIAL).
      DAD.
      ARG/LYS-RICH (BASIC).

```

DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF 2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003038; Link.
 DR InterPro: IPR003324; SGXSG.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR InterPro: IPR001304; lectin_c.
 DR Pfam: PF00008; EGF_1.
 DR Pfam: PF00059; lectin_c; 1.
 DR Pfam: PF02339; SGXSG; 62.
 DR Pfam: PF00084; sushi; 1.
 DR Pfam: PF00193; Xlink; 4.
 DR ProDom: PD000918; Link; 4.
 DR SMART: SM00032; CCP; 1.
 DR SMART: SM00034; CLECT; 1.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00409; Ig; 1.
 DR SMART: SM00445; Link; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS00290; IG_MHC; FALSE_NEG.
 DR PROSITE: PS01241; Link; 4.
 DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
 DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
 DR GlycoProfile: Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW EGF-like domain; Calcium; Alternative splicing; Repeat;
 KM Immunoglobulin domain.
 FT SIGNAL 1 16
 FT CHAIN 17 2364
 FT DOMAIN 17 2364
 FT DOMAIN 170 247
 FT DOMAIN 268 349
 FT DOMAIN 504 581
 FT DOMAIN 602 683
 FT DOMAIN 774 907
 FT DOMAIN 1433 2112
 FT DOMAIN 2113 2149
 FT DOMAIN 2114 2364
 FT DOMAIN 2161 2276
 FT DOMAIN 2280 2338
 FT DISULFID 51 133
 FT DISULFID 173 246
 FT DISULFID 199 220
 FT DISULFID 273 348
 FT DISULFID 297 318
 FT DISULFID 509 580
 FT DISULFID 533 554
 FT DISULFID 607 682
 FT DISULFID 631 652
 FT DISULFID 2117 2128
 FT DISULFID 2182 2274
 FT DISULFID 2250 2266
 FT DISULFID 2281 2324
 FT DISULFID 2310 2337
 FT CARBOHYD 126 126
 FT CARBOHYD 239 239
 FT CARBOHYD 333 333
 FT CARBOHYD 387 387
 FT CARBOHYD 611 611
 FT CARBOHYD 667 667
 FT VARSPLIT 2114 2150
 SQ SEQUENCE 2364 AA; 246359 MW; 6F83763420C3D4C CRC64;

Query Match 12.1%; Score 145.5; DB 1; Length 2364;
 Best Local Similarity 27.0%; Pred. No. 0.13;
 Matches 64; Conservative 27; Mismatches 71; Indels 75; Gaps 17;
 5 WPOSGLPIHPINKISTEPTGTD---RSVFC---QMEVLDRCYSIPGHRLLC--VSCI 55

DB 635 WLAGSLRPL--VTPRACGGDKPGVRYIYKIPNQGILD---PLSRHNAFCFRCVSA 689
 QY 56 -----KRAAGNPGPD-----PG-----PTSLPPPS-----TP 78
 DB 690 PSPEEEKSAFTAGPDVEWAVTQVGPVAAVPIGEETIIPGTVEPNKTEMLAYTP 749
 QY 79 GSPLPGQDDPADAAEPFGKPTG---SEDHOG-RATOLPGALDTSPPCTOHFAPEPTIPG 135
 DB 750 AGTLPLPGIP-----PTWPPGTEATEHTTGTGSPATVEVPSASEKRP--SEEPFPEEPF 802
 QY 136 ASWSISPTTGGPLPWGTQTPPTVPEDKQPGEDLRHGTSLPADLGRPPECPHT 192
 DB 803 ---SEKPPPPPEL-----FPSEKPPSEKPPSEB-----FPSEKPPPEELFPS 845
 RESULT 11
 PREP_HUMAN STANDARD; PRT; 331 AA.
 ID PREP_HUMAN
 AC P04280;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Salivary proline-rich protein precursor (Clones CP3, CP4 and CP5)
 DE [Contains: Basic peptide IB-6; Peptide P-H].
 GN PR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=85289325; PubMed=2993301;
 RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
 RT "Differential RNA splicing and post-translational cleavages in the
 RT human salivary proline-rich protein gene system."; J. Biol. Chem. 260:11123-11130(1985).
 RL J. Biol. Chem. 260:11123-11130(1985).
 RN [2]
 RP SEQUENCE OF 214-331.
 RA MEDLINE=86243355; PubMed=3521730;
 RA Kaufman D., Hofmann T., Bennick A., Keller P.;
 RT "Basic proline-rich proteins from human parotid saliva: complete
 RT covalent structures of proteins IB-1 and IB-6."; Biochemistry 25:2387-2392(1986).
 RL Biochemistry 25:2387-2392(1986).
 RN [3]
 RP SEQUENCE OF 276-331.
 RA MEDLINE=84161824; PubMed=6671974;
 RA Saitoh E., Isemura S., Sanada K.;
 RT "Further fractionation of basic proline-rich peptides from human
 RT parotid saliva and complete amino acid sequence of basic proline-rich
 RT peptide P-H."; J. Biochem. 94:1991-1997(1983).
 RL J. Biochem. 94:1991-1997(1983).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: K03204; AAA60185.1; -
 CC EMBL: K03205; AAA60186.1; -
 CC EMBL: K03206; AAA60187.1; -
 CC PIR: A03291; PIHUB6.
 CC PIR: C25372; C25372.
 CC MIM: 180989; -
 KW Repeat; Parotid gland; Multigene family; Saliva; Signal.
 FT SIGNAL 1 16
 FT CHAIN 214 331
 FT CHAIN 276 331
 FT VARIANT 106 238
 FT MISSING (IN CLONE CP-4).
 FT FTID-VAR_005561.

RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 [16]
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=93293968; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,
 RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakola P., Ryyanen M., Pearce W.H., Yao J.S.T.,
 RA Jackson K., Smilens V.V., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.E., Michaels S.N., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 RT in the triple-helical domain of type III procollagen are an
 RT infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 [17]
 RP VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 [18]
 RP VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 RT with aortic aneurysms.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 [19]
 RP VARIANT EDS-IV ARG-828.
 RX MEDLINE=94016385; PubMed=8411057;
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
 RT "The substitution of glycine 661 by arginine in type III collagen
 RT produces mutant molecules with different thermal stabilities and
 RT causes Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 30:690-693(1993).
 [20]
 RP VARIANT EDS-IV SER-957.
 RX MEDLINE=89109135; PubMed=2492273;
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine
 RT and causes Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:1349-1352(1989).
 [21]
 RP VARIANT EDS-IV VAL-960.
 RX MEDLINE=95266429; PubMed=7749417;
 RA Tromp G., de Paeppe A., Nuytlinck L., Madhathari S.L., Kuivaniemi H.;
 RT "Substitution of valine for glycine 793 in type III procollagen in
 RT Ehlers-Danlos syndrome type IV.";
 RL Hum. Mutat. 5:179-181(1995).
 [22]
 RP VARIANT EDS-IV GLU-1014.
 RX MEDLINE=92316511; PubMed=1352273;
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
 Pope F.M.;
 RT "A single base mutation in the gene for type III collagen (COL3A1)
 RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
 RT syndrome type IV. An unaffected family member is mosaic for the
 RT mutation.";
 RL Hum. Genet. 89:414-418(1992).
 [23]
 RP VARIANT EDS-IV ASP-1050.
 RX MEDLINE=90037070; PubMed=2808425;
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
 RT "Single base mutation in the type III procollagen gene that converts
 RT the codon for glycine 883 to aspartate in a mild variant of
 RT Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:19313-19317(1989).
 [24]
 RP VARIANT EDS-IV VAL-1077.
 RX MEDLINE=91374480; PubMed=1895316;

RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
 RA Pope F.M.;
 RT "Characterisation of a glycine to valine substitution at amino acid
 RT position 910 of the triple helical region of type III collagen in a
 RT patient with Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 28:458-463(1991).
 [25]
 RP VARIANT EDS-IV GLU-1173.
 RX MEDLINE=93022543; PubMed=1357232;
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
 Query Match 12.0%; Score 145; DB 1; Length 1466;
 Best Local Similarity 26.1%; Pred. No. 0.092;
 Matches 72; Conservative 19; Mismatches 71; Indels 114; Gaps 19;
 QY 2 EGQWYQSGPLHPINKISTEPTCTG-----DRSYF-----COMVYLDK----- 39
 DB 7 KQSWL-LIALHPITLILAQEAVEGCSHLQSYADRVWKPFCQICVDSGSVLCDDI 65
 QY 40 YC-----SIPGYHRLCCVSCIKKASGP-----NPGPDGPTSLP----- 73
 DB 66 ICDDDELCPNPEIP--FGECCAVCPQPTPTPTTPNGQPGQPKGDPGPGIGRNGDP 123
 QY 74 -----PFSTPGSPLP-----GPD-----PADAEPP 95
 DB 124 GIPGP-GSPGSPGPGICESCPTGPQNYSPQYDSYDKSGVAVGLAGYGPAGPPGP 182
 QY 96 GKPTGSEDHQHRATOLGALDTSRGTQHP-----PAPETPIYGSMSISPTPGGLP 149
 DB 183 G-PGTSGH-----PQS--PGSPGTGPPGEGQAGPSP--PGPAGIAGSPGAKD 230
 QY 150 WGTOTPPVPEDKQCPGEDLRHPTSLPADLPGRP 185
 DB 231 -GESGRPR--PGERGLP-----PGIKGPAGIPGP 260
 RESULT 13
 CAC3_HAEEO STANDARD: PRT; 295 AA.
 ID CAC3_HAEEO
 AC P16253;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cuticle collagen 3A3.
 GN 3A3.
 OS Haemonchus contortus (Barber pole worm).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
 OX NCBI_TaxID=6289;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90136718; PubMed=2615789;
 RA Shamsan L.M., Pratt D., Boisvenue R.J., Cox G.N.;
 RT "Cuticle collagen genes of Haemonchus contortus and Caenorhabditis
 RT elegans are highly conserved.";
 RL Mol. Biochem. Parasitol. 37:73-86(1989).
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 CC MISCELLANEOUS: THIS PROTEIN SHOWS A POTENTIAL TRIPLE-HELICAL
 CC REGIONS, WHICH CONTAIN GLYCINE AS EVERY THIRD AMINO ACID.
 CC -1- MISCELLANEOUS: IN ALL NEMATODE CUTICLE COLLAGENS, THE POLYPEPTIDE
 CC CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND
 CC CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND
 CC CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND
 CC -1- SIMILARITY: TO CAENORHABDITIS ELEGANS CUTICLE COLLAGENS.
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
DR EMBL: M32821; AAA29174.1; -
DR EMBL: M32820; AAA29173.1; -
DR InterPro: IPR002486; Col_cuticle_N.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 2.
DR Pfam: PF01484; Col_cuticle_N; 1.
DR Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
SQ SEQUENCE 295 AA; 29618 MW; 85355A431B6D3311 CRC64;

Query Match 11.8%; Score 142; DB 1; Length 295;
Best Local Similarity 34.3%; Pred. No. 0.037;
Matches 58; Conservative 7; Mismatches 50; Indels 54; Gaps 15;

51 CVCICKASGPNCPDPTSLP--PFTPGSP-LPG-----PQD-----P 88
91 CESCGR--PGP-PGP-PGPPGPRGP-GRPGAGLPVGLPPLPPGSCPEVSIIPCAACP 145
DB 89 ADAEPPGKP---TGSSEHOGHGRATQ--LPGALDTSPTGHPAPAPETP-IPGASWST- 140
146 AGPPGPGKPGPPGDPGDEGQPGPPGQDGLPGQGTGKP---PGPPGPGKPGSGSEVG 201
QY 141 -----SPTTPGGLPGKMTQTPTPVPRDKQGPGBDLRHGTSLLPADLPCR 184
DB 202 EDAECPEVAVPG-----DQGPPEPGPRGPGE---PGLQGPVGMPCQ 240

RESULT 14
CC12-CAEEL STANDARD; PRT; 316 AA.
AC P20630;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cuticle collagen 12 precursor.
GN COL-12 OR F15H10.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=90172409; PubMed=1689778;
PARK Y.-S., Kramer J.M.;
"Tandemly duplicated Caenorhabditis elegans collagen genes differ in
their modes of splicing";
J. Mol. Biol. 211:395-406(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Berks M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS. THE MATURE COL-12 AND COL-13 ARE IDENTICAL.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51622; CAA35954.1; -
DR EMBL: Z73972; CAA98257.1; -

DR PIR: S08169; S08169.
DR WormPep; F15H10.1; CE05638.
DR InterPro: IPR002486; Col_cuticle_N.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 2.
DR Pfam: PF01484; Col_cuticle_N; 1.
DR Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW Signal.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 316 CUTICLE COLLAGEN 12.
FT DOMAIN 128 157 TRIPLE-HELICAL REGION.
FT DOMAIN 176 202 TRIPLE-HELICAL REGION.
FT DOMAIN 206 235 TRIPLE-HELICAL REGION.
FT DOMAIN 240 266 TRIPLE-HELICAL REGION.
FT DOMAIN 269 304 TRIPLE-HELICAL REGION.
SQ SEQUENCE 316 AA; 30098 MW; 6CA81FF94706D42E CRC64;

Query Match 11.7%; Score 141.5; DB 1; Length 316;
Best Local Similarity 27.8%; Pred. No. 0.041;
Matches 58; Conservative 13; Mismatches 55; Indels 83; Gaps 13;

QY 51 CVCICKASGPNCPDPTSLP--PFTPGSP-LPG-----PQD-----P 70
DB 122 CCGSGGAGPAGPAGSPGQDAPGNDAPGAPGPGQDASHQDSCFCPCPPAGPEPS 181
QY 71 SLP-----PFTPGSP-----LPGPQDPAADAEPGKPTGSEDH-----OHGRATOLPG 114
DB 182 GAPGQKPGCAPGAPGAPGQSGALPGPPGAPGPPGAPGQ-PSNGNAGAPGAPGVVDPG 240
QY 115 ALDTSSPGTQHPAPAPETP-IPGASWSTPTTPGGLPGKMTQTPTPPE-DKQPPEDLRHP 173
DB 241 -----TPGAGP--PGSPGAG--APGQPGQA--GSSQPGGPPGQDAGAPGA---P 283
QY 174 GTSLPADLPGRRPP-----CHPTGT 194
DB 284 GAPGQAGAPGQDGGESGSCACDCHCPPRT 312

RESULT 15
CC13-CAEEL STANDARD; PRT; 316 AA.
AC P20631;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cuticle collagen 13 precursor.
GN COL-13 OR F15H10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=90172409; PubMed=1689778;
PARK Y.-S., Kramer J.M.;
"Tandemly duplicated Caenorhabditis elegans collagen genes differ in
their modes of splicing";
J. Mol. Biol. 211:395-406(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Berks M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS. THE MATURE COL-12 AND COL-13 ARE IDENTICAL.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X51623; CAA3595.1; -
CC DR EMBL: 273972; CAA98258.1; -
CC DR PIR: S08170; S08170.
CC DR WormPep: p15h10.2; CE05639.
CC DR InterPro: IPR002486; Col_cuticle_N.
CC DR InterPro: IPR000087; Collagen.
CC DR Pfam: PF01391; Collagen; 2.
CC DR Pfam: PF01484; Col_cuticle_N; 1.
CC DR Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
CC KW Signal.
CC KM CHAIN. 1 36 POTENTIAL.
CC FT DOMAIN 128 157 TRIPLE-HELICAL REGION.
CC FT DOMAIN 176 202 TRIPLE-HELICAL REGION.
CC FT DOMAIN 206 235 TRIPLE-HELICAL REGION.
CC FT DOMAIN 240 266 TRIPLE-HELICAL REGION.
CC FT DOMAIN 269 304 TRIPLE-HELICAL REGION.
CC FT DOMAIN 316 AA; 30100 MW; 00C6D08FBC4701AF CRC64;
CC SEQUENCE

```

```

Query Match 11.7%; Score 141.5; DB 1; Length 316;
Best Local Similarity 27.8%; Pred. No. 0.041;
Matches 58; Conservative 13; Mismatches 55; Indels 83; Gaps 13;

```

```

OY 51 CVCCKKASRP-----NPGPD-----PGPT 70
DB 122 CCSCGSAAPGASPGQDAGAPGNDGAPGAPGNGDASEDQTAGPDSFCFDCPAGPGRS 181
OY 71 SLR-----LPGFQDPAADAEPPKPTSEHD-----OHGRATQDLP 114
DB 182 GAGGQKGPSSAPGAPGQSGGALPGRPPGAPGPPGAGP--GSGNGAGAPGAPQGVVDVPG 240
OY 115 ALDTSSFGTGHPPAPETPIFGASWSISPTTGGLPWGTQTPPVPE--DKGQPGDELRRP 173
DB 241 ----TPGPPAGP--PGSPGAG-----APGQPGQA--GSSQPGGPGGAGAGA----P 283
OY 174 GTSLPNDLPGRPP-----CHPTGT 194
DB 284 GAPGQAGAPGQDSESGEGACHCHCPPPT 312

```

```

RESULT 16
CAEEL
OY 021184; 020807;
DB 01-NOV-1997 (Rel. 35, Created)
DB 01-NOV-1997 (Rel. 35, Last sequence update)
DB 01-NOV-1997 (Rel. 35, Last annotation update)
DB Putative cuticle collagen F55C10.3.
DB F55C10.3
DB Caenorhabditis elegans.
DB Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
DB Rhabditidae; Peloderinae; Caenorhabditis.
DB NCB1_TaxID=6239;
DB RN [1]
DB RP SEQUENCE FROM N.A.
DB RC STRAIN-BRISTOL N2;
DB RA White S.;
DB RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DB RN [2]
DB RP REVISIONS.
DB RC STRAIN-BRISTOL N2;
DB RA Jones S.J.M.;
DB RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

```

```

CC -----
CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z74036; CAA98487.1; -
CC DR WormPep: F55C10.3; CE11182.
CC DR InterPro: IPR000087; Collagen.
CC DR Pfam: PF01391; Collagen; 2.
CC DR Hypothetical protein; Cuticle; Connective tissue; Repeat;
CC KW Collagen; Multigene family.
CC KM CHAIN. 70 99 TRIPLE-HELICAL REGION.
CC FT DOMAIN 118 144 TRIPLE-HELICAL REGION.
CC FT DOMAIN 148 169 TRIPLE-HELICAL REGION.
CC FT DOMAIN 183 245 TRIPLE-HELICAL REGION.
CC FT DOMAIN 266 AA; 25616 MW; 98455680F1AAE22 CRC64;
CC SEQUENCE

```

```

Query Match 11.7%; Score 140.5; DB 1; Length 266;
Best Local Similarity 31.4%; Pred. No. 0.041;
Matches 53; Conservative 6; Mismatches 73; Indels 37; Gaps 9;

```

```

OY 50 CVCCKKASRP--NPGPD--PGPTSLPFGSPSPGPDPAADAE----- 93
DB 63 CDACCLPAPAGPAGTPGKGRGAPGAP--GLPGNPGRPPQPCDDPTTPPCQCPGQPP 120
OY 94 -PPKPTGSED-----HGRATQDLPALDTSSPG--TGHPPAPETP--IPGASWS 139
DB 121 GPPGPPGPGSGAGGNGNPGSPQDQDQ----PGAGGNKPGSPGNPAPGAPGQPGDAP 176
OY 140 ISPTTGGLPWGTQTPPVPEDEKGGPGEDELRRHPTSLPNDLPGRPP 188
DB 177 SEPTTPGAPGPGTTPGPPGQPGQPGQPGHD--GQFGAPGKPGNPNQGP 224

```

```

RESULT 17
CA44_HUMAN
ID CA44_HUMAN STANDARD; PRT; 1690 AA.
AC P53420;
DB 01-OCT-1996 (Rel. 34, Created)
DB 01-OCT-1996 (Rel. 34, Last sequence update)
DB 01-MAR-2002 (Rel. 41, Last annotation update)
DB Collagen alpha 4(IV) chain precursor.
DB COL4A4.
DB OS Homo sapiens (Human).
DB OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DB OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
DB OX NCB1_TaxID=9606;
DB RN [1]
DB RP SEQUENCE FROM N.A.
DB RC TISSUE=Kidney;
DB RX MEDLINE=95014445; PubMed=7523402;
DB RA Lehtonen A., Maruyama M., Mochizuki T., Tryggvason K., Reeders S.T.;
DB RT "Complete primary structure of the human type IV collagen alpha 4(IV)
DB chain. Comparison with structure and expression of the other alpha
DB (IV) chains."; 269:26172-26177(1994).
DB RL J. Biol. Chem. 269:26172-26177(1994).
DB RN [2]
DB RP SEQUENCE OF 1-23 FROM N.A.
DB RX MEDLINE=98196854; PubMed=9537506;
DB RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,

```


FT VARIANT 1030 1030 /FTID=VAR_008152.
 FT CUTICLE 1030 1030 G-> V (IN AS).
 FT VARIANT 1201 1201 /FTID=VAR_008153.
 FT VARIANT 1201 1201 G-> S (IN AS).
 FT VARIANT 1402 1402 /FTID=VAR_001913.
 FT VARIANT 1402 1402 P-> S.
 FT VARIANT 1572 1572 /FTID=VAR_008154.
 FT VARIANT 1572 1572 P-> L (IN AS).
 FT CONFLICT 1659 1660 /FTID=VAR_008155.
 FT CONFLICT 1659 1660 LQ-> FE (IN REF. 3).
 SO SEQUENCE 1690 AA; 164095 MW; E1E72F283A72BAAE CRC64;

Query Match 11.7%; Score 140.5; DB 1; Length 1690;
 Best Local Similarity 35.7%; Pred. No. 0.19;
 Matches 55; Conservative 2; Mismatches 56; Indels 41; Gaps 11;

QY 58 ASGPNPG-----PDGPTSLPFTSTPGSP-LPGPDADAAPEPKPTGSEDH 104
 DB 1237 SSGP-PGPGATGRARAKDIPDEGP-----PGDGGPPGPGPGARAPGPGLP-GSVDL 1286
 DB 105 QHGRATQLPGALDTSPTGTHPPAPETP-----IPGASWSISPTTPTG--PWGWTQPT 157
 DB 1287 LRGE---PG--DCGLPGPPGPPGPGPGYKGFPGCDKDGKGFPGPGGPGGPGFG 1340
 QY 158 PVEDKGPGEEDLRHPTGSLPADLGRPRPECHP 191
 DB 1341 P-PGKGLPGP---PGRKGPGLPGPRGPGPP 1369

RESULT 18
 CC34_CAEEL STANDARD; PRT; 298 AA.

AC P34687;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Cuticle collagen 34.
 GN COL-34.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=93013043; PubMed=1398138;
 RA Bird D.M.;
 RT "Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34
 gene 120:261-266(1992)."
 RL -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 LINKS.
 -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
 COLLAGENS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M80650; AAA27985.1; -
 CC PIR: JCI448; JCI448.
 CC InterPro: IPR002486; COL_cuticle.N.
 CC InterPro: IPR000087; Collagen.
 CC Pfam: PF01391; Collagen; 2.

DR Pfam: PF01484; COL_cuticle.N.1.
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
 FT DOMAIN 103 132 TRIPLE-HELICAL REGION.
 FT DOMAIN 151 177 TRIPLE-HELICAL REGION.
 FT DOMAIN 181 198 TRIPLE-HELICAL REGION.
 FT DOMAIN 215 277 TRIPLE-HELICAL REGION.
 SO SEQUENCE 298 AA; 29653 MW; 87E5370E0172D182 CRC64;

Query Match 11.5%; Score 139; DB 1; Length 298;
 Best Local Similarity 38.4%; Pred. No. 0.055; 58; Indels 22; Gaps 11;
 Matches 53; Conservative 5; Mismatches 58;

QY 60 GNPDPGPTSLPFTS-TPGSP-----LPGPDADAAPEPKPTGSEDHQHRATQ 111
 DB 151 GP-PGP-PGPPGPGDGGSGPSPGLPGQDAAPGEPGKGPGRPGAP-GAPEHQSECP 207
 QY 112 LPGAIDTSPTGTHPPAPETP-IPGASWSISPTTPTG--GLPWGWTQPTTPVEDKGPGEED 169
 DB 208 RGEPLRPGEPGPGEPAGPGGP-PG-----SPGPGADGSP-GDPGKPGNPGPD-GPGAD 259
 QY 170 LRHPTGSLPADLGRPRPE 187
 DB 260 -GNPAGPAGPPGSPGE 276

RESULT 19
 CCDC_CAEEL STANDARD; PRT; 302 AA.

AC P17657;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Cuticle collagen dpy-13.
 GN DPY-13 OR DPY-16 OR F30B5.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX WATERSTON R.;
 RA Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
 RL [1]
 RN VARIANT DPY13(E225).
 RP MEDLINE=93013043; PubMed=1398138;
 RA Bird D.M.;
 RT "Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34
 gene 120:261-266(1992)."
 RL -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 LINKS.
 -1- DISEASE: MUTATIONS IN DPY-13 AFFECTS THE BODY SHAPE.
 -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
 COLLAGENS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 20:09:28 ; Search time 87.84 Seconds

(without alignments)
2465.729 Million cell updates/sec

Title: US-09-938-330-20

Perfect score: 6954

Sequence: 1 MAPLRALLSYLLPLHLCALCT.....PTGFTLLCVLRDLSQRLRHT 12532

Gap table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phase:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_rviro:*
16: SP_bacteriophage:*
17: SP_archaeophage:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1072.5	15.4	2165	5	Q19791
2	1038.5	14.9	1054	5	Q9W493
3	962	13.8	1229	5	Q9V661
4	948	13.6	1427	4	Q96L37
5	810	11.6	364	4	Q96AY5
6	774.5	11.1	550	4	Q9NSJ8
7	707	10.2	2174	5	Q9G0R0
8	707	10.2	3060	5	Q9YAV4
9	699.5	10.1	2167	5	Q76840
10	683.5	9.8	1235	4	Q95428
11	681.5	9.8	1280	11	Q9EPX2
12	677	9.7	1444	5	Q17591
13	668.5	9.6	3198	5	Q9U8G8
14	663.5	9.5	1572	5	Q44938
15	622	8.9	951	4	Q60345
16	614	8.8	1062	5	Q19204

17	597.5	8.6	525	4	Q96RW4	Q96RW4 homo sapien
18	595.5	8.6	766	4	P82987	P82987 homo sapien
19	556.5	8.0	860	5	Q22580	Q22580 caenorhabdi
20	518	7.4	1014	5	Q95R33	Q95R33 drosophila
21	515	7.4	1091	5	Q9W126	Q9W126 drosophila
22	480.5	6.9	269	6	Q9GL54	Q9GL54 oryctolagus
23	433	6.2	898	4	Q9F724	Q9F724 homo sapien
24	430	6.2	1059	5	P90884	P90884 caenorhabdi
25	427	6.1	364	4	Q9UGQ1	Q9UGQ1 homo sapien
26	388.5	5.6	1637	6	Q9XSV8	Q9XSV8 bos taurus
27	384.5	5.5	610	13	Q93523	Q93523 bothrops ja
28	379	5.5	227	4	Q9BX28	Q9BX28 homo sapien
29	379	5.5	340	11	Q91256	Q91256 mus musculu
30	362.5	5.2	610	13	Q9120	Q9120 agkistrodon
31	361	5.2	609	13	Q90282	Q90282 crotalus at
32	353	5.1	728	4	Q9BZ11	Q9BZ11 homo sapien
33	350.5	5.0	873	13	Q42595	Q42595 xenopus lae
34	342.5	4.9	918	4	Q9H013	Q9H013 homo sapien
35	342	4.9	918	4	Q9BZ15	Q9BZ15 homo sapien
36	341	4.9	920	11	Q35674	Q35674 mus musculu
37	339.5	4.9	616	13	Q90495	Q90495 echis carin
38	337	4.8	609	13	Q9W6M5	Q9W6M5 agkistrodon
39	332	4.8	1023	5	Q967H9	Q967H9 strongyloce
40	325	4.7	857	4	Q9UHP2	Q9UHP2 homo sapien
41	323.5	4.7	4123	4	Q75851	Q75851 homo sapien
42	320.5	4.6	1569	5	Q9VFN0	Q9VFN0 drosophila
43	319	4.6	1081	5	Q90631	Q90631 drosophila
44	318.5	4.6	5636	4	Q96RW7	Q96RW7 homo sapien
45	317.5	4.6	622	4	Q9H8X0	Q9H8X0 homo sapien

ALIGNMENTS

RESULT 1
ID Q19791 PRELIMINARY: PRT: 2165 AA.
AC 01-NOV-1996 (TRIMBLrel. 01, Created)
DT 01-NOV-1996 (TRIMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TRIMBLrel. 19, Last annotation update)
DE F25H8.3 PROTEIN.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditiida;
OC Rhabditiida; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadasy S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Haynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gattner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sherry-Mieg J., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkerson-Sprat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
ellegans.";
RL Nature 368:32-38(1994).
DR EMBL: Z69361; CAAG3288.1;
DR EMBL: Z69360; CAAG3288.1;
DR EMBL: Z69360; CAAG3287.1;
DR EMBL: Z69360; CAAG3287.1;
DR HSSP: P15167; IDTH.
DR MEROPS: M12.135; -.

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markov G., Milshina N.V., Mobarry C., Morris J., Mosnig D.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon S., Pollard J., Puri V., Reese M.G.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 EMBL: AE003709; AF55199.1;
 MEROPS; M12.0PB;
 FLYbase; FBgn0038340; CG6107.
 InterPro: IPR002870; Pep_M12B_propep.
 DR InterPro: IPR001590; Repolysin.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR000130; Zn_M12B_propep.
 DR Pfam: PF01562; Pep_M12B_propep; 1.
 DR Pfam: PF01421; Repolysin; 1.
 DR Pfam: PF00090; TSP_1; 2.
 DR SMART; SM00209; TSP1; 3.
 DR PROSITE; PS50215; ADAM_MERPRO; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR HydroLase; Metalloprotease; Zinc; C1293180AB19B700 CRC64;
 KW SEQUENCE 1229 AA; 138136 MW; C1293180AB19B700 CRC64;
 SQ
 Query Match 13.8%; Score 962; DB 5; Length 1229;
 Best Local Similarity 27.9%; Pred. No. 1,6e-66;
 Matches 299; Conservative 142; Mismatches 412; Indels 218; Gaps 45;

QY 532 GKWCFKGHCIMKSPBQTYGO--DGGSSWTKFGSCRSRGGVRSRSHSNPNPSPAYGR 589
 DB 619 GWCRCRGKCV--SNHGGROYNGMGWIPPTPCSLTCGGGVDSEKRCQPPENGR 676
 QY 590 PCLGPMFEYOYCNSECEPTEDEPFAOOCAR--RNSYVHQAQKSHVWP-YEPDDAOK 645
 DB 677 YTGSRKKRYKSCNTGCPFGSMDPREQCYAMNGRNNMIPGVNPDTKVVPKYEK--A 732
 QY 646 CELTICQADTGCVYVMMNOYVHDGTCSRDYISVACARECVPGGDKRVGSMKMDKCGV 705
 DB 733 CILFRMDKRYTYEMLKSNVTDGTSCAV--DSFDCVNICRPGAGDNLNIAKIDKCGV 791
 QY 706 CGDNSHCFTYKGTGKASKQAG-----ALKVQIPAGAHQIIEALERSPHRSV 755
 DB 792 CGGRNDTCHVYGNL-LVSNLLGLNDGNENPKTLTYVRIRPKGASNIITQRGYDDQFI 850
 QY 756 VKNQYTGSEILNPKGKEATSRFTAMGLEWE-----DAVEDAKE--SLKTGSPLEALAI 808
 DB 851 VLTDRDNELLNGKFLKTYPLKFFVAGVMOYTGSVSEVQVNTTYSKLSRDLIVQIIS 910
 QY 809 LALPTEGGPRRSILAKYVYIHEDILPLIGSNVLLIEMDTYEMALKSNAPCSKACGGIQ 868
 DB 911 LDVSPKRODTYVLTSTYIT--DKPPDYEA-----EVEITYEMQADSNCDISLCEG--- 959
 QY 869 FTKYGCRRRDHMM-----VQRHLCDBKRRPPIRRRCNQHPSCOPVWTEMG 917
 DB 960 -----RSHRLPACISTTGQVKAPOFCDSKAMPKIDRACNT--DCRLNLIVT--SIS 1008
 QY 918 ACSRSCKGLGVTGRTGICLLPLSN-GTHKVMARKC-----AGDPEARPCPLARPCAO 971
 DB 1009 ECSNACGELGTREKYACVQYTFNMORSNIYMSYCKLKFVAYHECEBEG----- 1060
 QY 972 WRIGAMSCSATCGEGIQOQVYCFRTNANSLGHCGRDPT-----VOYCSLPACGNN 1024
 DB 1061 WYLSWSTSCSGTSGTSGQREAHCHYLH-NSRVSDDLCNRTMPLHLTLIGINTESCPTY 1119
 QY 1025 HQN-STVRADVWELGTPEGQWVPOSGPLHPINKISSMCAEPTCTDRSYFC 1074
 DB 1120 TKSPNALAVSNVYI-----GEC-NEWCEKTRTSVSC 1151
 RESULT 4
 ID 096L37 PRELIMINARY; PRT; 1427 AA.
 AC 096L37;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE VON WILLEBRAND FACTOR-CLEAVING PROTEASE PRECURSOR.
 GN ADAMTS13.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX PubMed=11557746;
 RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,
 RA Fujikawa K.;
 RT "Structure of von Willebrand factor-cleaving protease (ADAMTS13), a
 RT Metalloprotease involved in thrombotic thrombocytopenic purpura.";
 RL J. Biol. Chem. 276:41059-41063(2001).
 DR EMBL: AY053376; M117652.1;
 KW Signal; Protease.
 FT SIGNAL 1 33 POTENTIAL.
 FT CHAIN 75 1427 VON WILLEBRAND FACTOR-CLEAVING PROTEASE.
 SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABCI4442 CRC64;

Query Match 13.6%; Score 948; DB 4; Length 1427;
 Best Local Similarity 28.4%; Pred. No. 2.5e-65;

Matches 284; Conservative 145; Mismatches 400; Indels 172; Gaps 44;

```

QY 249 RRRRAKAGSYIEVLVVDSDYVRFRKKEVQNYVLTLMNYDEIYHDSGLVHNIAL 308
Db 70 QORRAAGGIHLLELVAVGPDVFOAH-QDETERXYVLTNINIGELLRLDPSLGAQFRVHL 128
QY 309 VLLIHWGYROSILIERENPFRSLQYCRMAHSQORODPSHAHHHVVLTLPDQDG-PS 367
Db 129 VKMVLTEPECAPNI-TANLTSLSLSCGMSQOTINPEDDTDPGHADLVLTITPDLPL 187
QY 368 GYAPYTG-----CHPLRSCALNHEDGFSSAFVIAHETGHVLMEDHG-OGNGCADETS 420
Db 188 GNRQYRGVYTOLGACSPFTMSCLITEDTGFDIGVIAHEIGHSPGLEHIDGAPSGGCP-- 245
QY 421 LGSVAFLVQAAFRHFRMSRCKLSRLYPS--YDCLLDPPDPAMPQ-----PPEL- 471
Db 246 -GHWASAGAAPRAGLAMSPCSRKQLSLSLAGARCV---WDPFRPQSGAGHPDDAQ 300
QY 472 PGINYSMEQCRFDGSGSYQTC-LAFRTPECKQIMC-SHPDNPYCKTKKGPPLDGTG 529
Db 301 PGLYASANEQCRVAFPKAVACTFAREHDMCQALSCHTPDLPOSSCRLLVPLLDGTG 360
QY 530 APGWKCFRGHC---IMKSPEDTIGDGGSSWTKFGSCSRSGGCVKRSRSCNNPSPAY 586
Db 361 GVEKWSCKGRCSRLVELTP--IAAVHGRMSSWGPFRSPCSRSGCGGVYTRRRQNNRPAF 418
QY 587 GGRPCLGPMFEYQVCSNEECPTGYEDFRAQOQAKRN-----SYVHQNAKHSW 635
Db 419 GGRACVADLQAEKNTQACEKTOLEFMSQOAKRTDQPLRSSPGGASFF-HMGAA--V 474
QY 636 PEPDDAOKCELLQSGADTGVVFMNVVHDTGTSYRDP-----YSVARGECYVPGC 690
Db 475 PHSQGD--ALCRHMCRAIGSEFIKRGDSFLDGTROMPSPREDTGLSLCYSVSSCTETG 532
QY 691 DKEVSKAKDKCGVCGGDSHGRTYKG--TLKASKQAGALKVIOIPAGARIHQLEALE 748
Db 533 DGRMSQVWDRQVCGGDSNSTSPRGSPTARAREYVTLVTP-----NLTSY 584
QY 749 KSPHRSVYKN---QVTSFILNPKGEATSRPTAM-----GLEWEDAVEADK---ESLK 797
Db 585 IANHRPLFTHLAVRIGRYVAVAGKMSISPTTYPSLLEGGRYKVALIEDRLPLEETR 644
QY 798 TSGPLPE-AIALLALPTEG--PRSSLAAYKVIHEDLLPLGNNVLEMDTYEAL 853
Db 645 IWGLQEDADIQYRRYGEYGNLTPDITFTY-----FQKPRQAWY- 687
QY 854 KSWA-----PCSKACGGGIOFTYGCRRRRDHNVYRHLCDHKRPRIRRCNQHPCSP 909
Db 688 --MAAVGPGSVSCAGALRWVNSCLDOAKKELVETVOCGSOQPPAMPDEACVLEPC-PP 744
QY 910 VWTEWAGCSRSCKLGVQTRGICLLPLNSGTHKVMKAKACGDRPARP-----C 963
Db 745 YMAVGDEGPGCSASCGG-GLREPRVCVEAOGSLTKLTPPARBCRG---AQQPAVALLGC 799
QY 964 LRVPCPAOMLIGMOSCSATCGEGIOOROVYCHTANSLGHEGDRPDY-----QVC 1016
Db 800 NPQPCPAKMWVSPSSCTSGAGGLALENETCVPGADGLEAVTRPGSVDEKLPAPEPC 859
QY 1017 SLPAC--GGNHQNSTVRADWELGTPEGO-----WVPSQGLH----- 1052
Db 860 VGMSCPGMWHLDATSGE-KAPSPWGSIRTGAAAHWTVVAASCSYSCSGRIGMELRF 917
QY 1053 -----PINKISSMCAAEPCGTGRSVFCQMEVLDRYCSIPGYHRL--CCVSC---I 1097
Db 918 LCMDSALRYVDE--ELCGIASKPGSRREVQAVP---CPARMQYKLIACSVSCGRGV 971
QY 1098 KKA--SGPNVGPDPPTSLPPTSTGSLPGGODPADAAEP 1136
Db 972 RRLIYKARAHGEDDEEILLDIOCGGLPRPEDE-ACSLER 1011

```

RESULT
Q96AYS

5

```

ID Q96AYS PRELIMINARY; PRT; 364 AA.
AC Q96AYS;
DT 01-DEC-2001 (TREMblrel, 19, Created)
DT 01-DEC-2001 (TREMblrel, 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel, 19, Last annotation update)
DE HYPOTHETICAL 40.2 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA (MEMO CELL LINE);
RA Blum H., Bauersachs S., Mewes H.W., Well B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162080; CAB82413.1;
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR001128; Cyt_P450.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF0090; tsp_1; 3.

Query Match 11.6%; Score 810; DB 4; Length 364;
Best Local Similarity 43.6%; Pred. No. 2,4e-55;
Matches 157; Conservative 43; Mismatches 76; Indels 84; Gaps 9;

QY 844 EEMDTYEMALKSNAPSCAKAGGIGFTYGCRRRRDHNVYRHLCDHKRPRIRRCNQ 903
Db 1 EELDTFEMALKSNWQSCPCGGGFQYTKYGCRRKSNKMHRSFCBANKPKPIRMWNI 60
QY 904 HPCSPVWVTEWAGCSRSCKLGVQTRGICLLPLNSGTHKVMKAKACGDRPARP 963
Db 61 QECHEPLIMVADEWHCKTKTCGSSGYOLRTYKCLPLDGTNNSVHSKTCMGPRPSRRPC 120
QY 964 LRVPCPAOMLIGMOSCSATCGEGIOOROVYCHTANSLGHEGDRPDYVQVSLPACG 1023
Db 121 NRVPCPAOMLIGMOSCSATCGEGIOOROVYCHTANSLGHEGDRPDYVQVSLPACG 1023
QY 1024 NHQNSTVRADWELGTPEGOVPSGRLPIKISSMCAAEPCGTGRSVFCQMEVLDRYC 1083
Db 177 -----EPLDQKSLFCQMEVLAHYC 196
QY 1084 SIPTYHRLCCVSCIKKASG-----PNEGPPG-----PTSLPPE--STP 1120
Db 197 SIPTYHRLCCVSCIKKASG-----PNEGPPG-----PTSLPPE--STP 1120
QY 1121 G-----SPLGQPDADAAEPCKPTGSEDDHGGRATQLPGALDT-----SSPQT 1166
Db 257 AKKMSLSSISVGGPNAYA-AFRPNKSPDGANLRO--RSADQAGSKTVRLVTPSSPTK 313

RESULT 6
Q9NSJ8 PRELIMINARY; PRT; 550 AA.
ID Q9NSJ8;
DT 01-OCT-2000 (TREMblrel, 15, Created)
DT 01-OCT-2000 (TREMblrel, 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, Last annotation update)
DE HYPOTHETICAL 60.3 KDA PROTEIN (FRAGMENT).
GN DKEZP762C1110.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA (MEMO CELL LINE);
RA Blum H., Bauersachs S., Mewes H.W., Well B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162080; CAB82413.1;
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR001128; Cyt_P450.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF0090; tsp_1; 3.

```


QY 1047 QSGFLHPINKISSMCAAEPCGDRSVFCOMEV--LDRCYSPGYHRLCCVSCITKASGPN 1104
 Db 557 EE-----FCEPEVPTLSRPPCKSPKCEAQMVSSEMSKASAPC 592
 QY 1105 P-----GPDGPTSPSPSTPSPGPPGPDADAAEPKPGTSGEDHQ--HGRAT 1152
 Db 593 GKGVKSRIVICGEFDKXT-----VTPADDDSKNKTKTKTESSEDDCEGEK 637
 QY 1153 QLPALDTS-----SPTQHPFAPE--TPPGASMSISPTPG 1188
 Db 638 VCPGEWFTGPGKCKSPGCGGERVREYLCLSNCTKSVNCDKEEVLSEKNSACTEDE 697
 QY 1189 GLPWGWTQTPPYEDKGGEDLRHGTSLPAD 1222
 698 ILP--LSTDKPTEDDE---ECCDEGDIETLSD 725
 RESULT 8
 Q9VAV4 PRELIMINARY: PRT: 3060 AA.
 ID Q9VAV4: Q9VAV3:
 AC Q9VAV4: Q9VAV3: 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 16, last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, last annotation update)
 DE CG1540 PROTEIN.
 GN PPN OR CG1540 OR CG18436.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Phyloroidae; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132.
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Sulten R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
 Burris K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.D., Wei M.-H., Idegawa C.,
 Jafarli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostnefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon S., Nusskern D.R., Paclik J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Splet E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL: AE003765; AAF56794.2; -;
 DR EMBL: AE003765; AAF56795.2; -;
 DR HSP: P12111; 2KNT.
 DR FLYBase: FBgn0003137; Ppn.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003598; 19_c2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00047; 19; 3.
 DR Pfam: PF00014; Kunitz_BPTI; 12.
 DR Pfam: PF00090; tsp_1; 5.
 DR Pfam: PF00095; Wap; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00408; ICG2; 3.
 DR SMART: SM00131; K0; 12.
 DR SMART: SM00209; TSP1; 7.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4 DISULFIDE CORE; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 11.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 12.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS00092; TSP1; 3.
 KW Alternative splicing; Immunoglobulin domain;
 KW Serine protease inhibitor.
 FT VARSPPLIC 2803 2803 L->SYSPV (IN SHORT ISOFORM).
 FT VARSPPLIC 2844 2854 ENFKRMEDSGI->VASPPLHPNAV (IN SHORT ISOFORM).
 FT VARSPPLIC 2855 3060 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 3060 AA; 331579 MM; ACA31D3EE558C7C0 CRC64;
 Query Match 10.2%; Score 707; DB 5; Length 3060;
 Best local Similarity 29.2%; Pred. No. 6.8e-46;
 Matches 220; Conservative 75; Mismatches 311; Indels 148; Gaps 29;
 QY 545 PEQ--TYGGDGG---WSSMTKFGSCSRSCGGGVSRSRSCNNPSPAYGAPRCIPMPREY 598
 Db 4 PESSVTPGEGNDPDEMTWSPSCSRSCGCGVSYOTRECLRRDR-GEAVCGGSGRRY 62
 QY 599 QVNSSECEGTEDEDRAQOC--RNSYVHONAKHSVPEPPDDAKCELTCQSDATG 656
 Db 63 FSCNTQDCEESDRAQOCSEFRDQF---DGYVEWVY--TNAPPCELNCPGGER 117
 QY 657 DVVFNQVVDHGTGCRSPYRPSVCAAGCEVPYCGDKKESMKKADKCGVCGDMSHCRVY 716
 Db 118 FYRQREKVVDDTRCKND-LDYCVNGECMTVYGDMMIGSDAKEDKCKCGGDSSTCKTI 176
 QY 717 KTLGKASKQAGALKLVQIPAGARHIQLEALEKSPHSRVKNQVTSFILNPKKEATSR 776
 Db 177 RMTITTKDLAPYNDLLLPBEGATNIRIEFVPSNVYLAGCNH-SGHYYLNGMDRIDPR 235
 QY 777 TTTAGCLEHDAVED-----AKSLKTSPLDEALAILALPTBEGSPSSATKYVHEDL 832
 Db 236 PWFANSMWNTQKRMGEFAAPDOLTCGSISESLEIYVLQEK---NISDIYESIPESL 292
 QY 833 LPLIGSNVNLLEEMQTYMALKSNAPCSKAGGSIQFTYKGCRRRRHHNVQHLCDHK 892
 Db 293 SH-----SQDTHVTHQFNACASGSGSQRKVTNNRITLAEVNSLDDKS 343
 QY 893 RPKPIRRRCNHPSPQPVWTEEMGACSRSGKIGVOTRGICLLPLTSNTHKVMPAKAC 952
 Db 344 KPYE-EQACGTEPCA-PRHWEGEMSKSGSGDPGNRBITERISSSEHNYVEEDAYC 401
 QY 953 ---AGDRPARRPCLR--VCPDAQMRIGAMSQCATCBESIQGRQVAVCRTNANSLGH--- 1004
 Db 402 LKEVGNKPKATQKCCNRPVKNCP-KYHLGPMTPDKILCGDKQTRKYTCFTEEN--CHKRV 458
 QY 1005 -----CEGDRDVTQVQVSLPACGNH-----GNSTVRADVVELGTPBEGQWP 1046

KW Hypothetical protein; Immunoglobulin domain;
KW Serine protease inhibitor.
SQ SEQUENCE 1235 AA; 133477 MW; A0B44CCE4F38E350 CRC64;

Query Match 9.8%; Score 683.5; DB 4; Length 1235;
Best Local Similarity 29.8%; Pred. No. 1.3e-44;
Matches 214; Conservative 71; Mismatches 283; Indels 149; Gaps 36;

QY 530 AFGKMCFGHCIMKSPEDQTYGGDGWSSMTKFGSCSRSCGGVRSRSCNNPSPATYGR 589
DB 14 AFG-----SSAPYVRROSDTWGPMWSQWSPCSRTCGGVSRFRPCYS-QRRDGS 62
QY 590 PCLGPMFEYQVQCNSECPGYEDFRAQOCARNSYVHQNAKHSWVPYEPDDAQKCELI 649
63 SCVGFARSHRSCRTSCPDGARDFRAEQAIEDG-AEFGRRYRWLPY--SAPKCELN 119
650 COSADTGQVFMNQVYHGTIC--SYRDPYSVCARGECVPVGCDEKVGSMKADKCGVCG 707
DB 120 CIPKGENFYKHKRAVVDGTCEPGKRD--VCYDGSQVVGCDHELDSSKQEDKCLRCG 176
QY 708 GDNSHCRTVKGTLGKASKQAGALKVQIPAGARHTQIEALEKSPHRSYVKNQVTSFTLN 767
DB 177 GDGTCYVAGTFDANDLSRGYNQILLVPMGATSTLIDEAASRNFVAVKN-VRGEYTLN 235
QY 768 PKGEATSRFTTAMG--LEMEDAVED--AKESLKTSGPLPAIATIALPTEGGPRSLA 823
DB 236 GHWTEARALPAASTILHYERGAEGDLAPERLHARGTSEPIVIELISQE--PNPGVH 292
QY 824 YKYVHEHLLPLIGSNVNLLEMDTYEALKSWAPCSKACGGGTOFTKYCCRRRRDHMV 883
DB 293 YEY--H--LP-----LRRPSPGFWSHGSWSDSAECGGHQSRLVFC--TIDHEAY 338
QY 884 QRHLCDHKRPKPIRRRCNOHPCSP-----VWVTEEMGACS 920
DB 339 PDHMQQRPRPAD--RSCNLHPCEPTKRTSTLHPRGAMRLAGAORVCGNSWKAGPMAPCS 397
QY 921 RSCGKLVQVOTGICOLLPLNSGTHKVMPPAKACAG--DRPEARPPCLRVPCPAQWRLGWS 978
DB 398 ASCGG-GSOSRSYVCISSDGAGIOEAVEAEACAGLPGKPAIQACNLQRC-AAMSPEPWG 455
QY 979 QCSATCGEGIOOROVORTNANSIGHCEG---DRPTVOYCSLPACGNGHONSTVRAD- 1033
DB 456 ECSVSCGVGKRSVTCGERGSLHTAACSLEDRPPLTEPCV-----HEDCPILSDQ 508
1034 VMEIGTEGQWVPOSGPLHPINKISSMCAEPCTGDSVFCOMEVILDRYCSI-PGYHRLC 1092
509 AMHYGT-----W-----GLCSKSSSGTR---KROYI---CALGPPSH--- 540
QY 1093 CVSCITKASGP-----NPGDDPGPTSLP-----PFSTPGSPPLPGPD-PADAAEPG 1138
DB 541 CGSL--QHSKRVDEPCNTQPCILPQLEVPMQDVHTPASNPWML-GQESPASAPIPA 597
QY 1139 KPTSEDHQHGATQLPGLDTSPTQHPFAETPTPGASW--SISPTTGGGLPWG 1193
DB 598 TP-----AVGLRAPRLQTSRVLV---RMPHGISRASVARIPLWG 634

Search completed: August 28, 2002, 20:14:02
Job time: 274 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 00:04:09 ; Search time 88.81 Seconds

(without alignments)
409,064 Million cell updates/sec

Title: US-09-938-330-22_COPY_1040_1249
Perfect score: 1206
Sequence: 1 PEGQWVPSGSLPINKISS.....PTGFTLCVLPDSQLAGHT 210

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:
2: SP_BACTERIA:
3: SP_FUNGI:
4: SP_HUMAN:
5: SP_INVERTEBRATE:
6: SP_MAMMAL:
7: SP_MNC:
8: SP_ORGANELLE:
9: SP_PHAGE:
10: SP_PLANT:
11: SP RODENT:
12: SP_VIRUS:
13: SP_VERTEBRATE:
14: SP_UNCLASSIFIED:
15: SP_RVIRUS:
16: SP_BACTERIAP:
17: SP_ARCHAEP:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	204	16.9	364	4	096AY5	096AY5 homo sapien
2	180.5	15.0	839	10	09SN46	09SN46 arabidopsis
3	175.5	14.6	585	12	041935	041935 murid herpe
4	163	13.5	345	5	095W2	095W2 meloidogyne
5	160	13.3	319	5	017038	017038 caenorhabdi
6	160	13.3	724	11	099LH3	099LH3 mus muscui
7	158	13.1	651	4	09NX19	09NX19 homo sapien
8	158	13.1	845	4	09HB5	09HB5 homo sapien
9	158	13.1	845	4	09HAU3	09HAU3 homo sapien
10	156.5	13.0	309	5	018751	018751 caenorhabdi
11	156	12.9	845	4	09HB8	09HB8 mus muscui
12	155.5	12.9	768	11	099N39	099N39 mus muscui
13	154	12.8	577	10	09STP1	09STP1 arabidopsis
14	153	12.7	408	12	09J2H0	09J2H0 macaca mula
15	151.5	12.6	305	5	025467	025467 meloidogyne
16	151	12.5	448	12	09WRM2	09WRM2 macaca mula

17	150.5	12.5	316	5	093207	093207 caenorhabdi
18	149.5	12.4	301	5	019763	019763 caenorhabdi
19	149.5	12.4	995	4	09Y2M4	09Y2M4 homo sapien
20	149.5	12.4	1012	4	043393	043393 homo sapien
21	149.5	12.4	1012	4	075359	075359 homo sapien
22	149.5	12.4	1268	4	075046	075046 homo sapien
23	149.5	12.4	1566	4	09P2R6	09P2R6 homo sapien
24	149	12.3	325	5	017036	017036 caenorhabdi
25	148.5	12.3	555	10	09EP06	09EP06 chlamydomon
26	148	12.3	676	6	095J79	095J79 sus scrofa
27	146.5	12.1	846	11	099N40	099N40 mus muscui
28	146	12.1	230	5	022260	022260 caenorhabdi
29	145.5	12.1	299	5	019218	019218 caenorhabdi
30	145.5	12.1	316	5	093210	093210 caenorhabdi
31	145.5	12.1	330	10	049870	049870 hordium vul
32	145.5	12.1	566	6	095J71	095J71 sus scrofa
33	145	12.0	297	4	016038	016038 homo sapien
34	145	12.0	297	5	023263	023263 caenorhabdi
35	145	12.0	299	5	020087	020087 homo sapien
36	145	12.0	358	4	008805	008805 drosophila
37	145	12.0	1180	5	09VRM2	09VRM2 drosophila
38	144.5	12.0	313	5	020922	020922 caenorhabdi
39	144.5	12.0	356	5	020921	020921 caenorhabdi
40	144.5	12.0	511	6	095J70	095J70 sus scrofa
41	144.5	12.0	888	13	090796	090796 gallus gall
42	144	11.9	362	5	09VC23	09VC23 drosophila
43	143.5	11.9	196	5	09GCT8	09GCT8 heterodera
44	143.5	11.9	642	10	09LD34	09LD34 cryptocodi
45	143.5	11.9	862	11	09JIK1	09JIK1 rattus norv

ALIGNMENTS

RESULT 1
096AY5 PRELIMINARY: PRT: 364 AA.
ID 096AY5
AC 096AY5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 40.2 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA, AND CHORIOCARCINOMA;
RA Strausberg R.;
Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016451; AA016451.1;
KW Hypothetical protein.
FT NON_TER
FT 1
SQ SEQUENCE 364 AA; 40210 MW; A3D54D967D67FD30 CRC64;

Query Match 16.9%; Score 204; DB 4; Length 364;
Best Local Similarity 38.7%; Pred. No. 2.5e-08;
Matches 58; Conservative 12; Mismatches 36; Indels 44; Gaps 8;

0Y 12 LHPINKISTEPCDDRSVFCQMEVLDYRCSIPGHRLCVCYCIKKASG-----60
DB 171 LPPCN-----DEPCLDCKSTFCQMEVLAAYCSIPGYNKLCSCSCSRSTLPPPYLAEAE 226
0Y 61 -----PNEGPPPG-----PTSLPPF--STPG-----SPLPQOPADAAEPGKPTG 100
DB 227 TRDDVYNSPDLPRSLVWPTSLVPRHSETPAKKMSLSISVGGPNAYA-AFRPNKSPDS 285
0Y 101 SEDRHGGRATQPGALDT-----SSPETQ 124
DB 286 ANLRQ--RSAQAGSKTVRLVTPSPSPPTK 313

Matches 72; Conservative 14; Mismatches 91; Indels 82; Gaps 14;

OY 3 GOWP-----OSGPLPINKISTEP---CTGDRSVFCOMEVLDRCYCSIRKASGP 52
DB 92 GOWVENEYEAQOTGPSTVSAISQPSGATYQCAAGTQVYAPVAPV-----CC 144
OY 53 SCIRKASGP-----NPGPD-----PGPTSLPFPSTGSLPFGPDP 88
DB 145 TCHGSPGPPIGEGEPGPGEDGPNKGDTSGKDARILPAPLEPPLICPPGP-AGPQGP 203
OY 89 ADAEPPCK-----PTGSEDHCHGR-ATQLGALDTSSTPGQHP 127
DB 204 AGAGGPPGSLCEPPKDGVPBGVQVGHGPPRGREGPRGAGSPGRL-TPVPGQGA 262
128 APETPIPGASWSISPTPG--GLPMWGTOTPTTPVEDKGGQGEDLHNGTSLPADLPGRP 185
263 GP-----PGVVG--PPAGAGAAAGPGGSGFEGPPRPGRPGNPGREGP-PGAPPGAPPGPD 315
OY 186 -----PPCHPGTGF 195
DB 316 GEKGSCEHCEPRTPPGYF 334

RESULT 5
ID 017038 PRELIMINARY; PRT; 319 AA.
AC 017038;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 31.3 KDA PROTEIN.
GN T15B7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA MEDLINE=9069613; PubMed=9851916;
RX None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Pauley A., Gattung S.;
RT "The sequence of C. elegans cosmid T15B7.";
RT Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF022985; AAB69959.1; -;
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR002486; COL-cuticle_N.
DR Pfam: PF01391; Collagen_2.
DR Pfam: PF01484; COL-cuticle_N; 1.
KW Hypothetical protein.
SQ SEQUENCE 319 AA; 31292 MW; 5242AAEC668AB488 CRC64;

Query Match 13.3%; Score 160; DB 5; Length 319;
Best Local Similarity 31.6%; Pred. No. 5.3e-05;
Matches 56; Conservative 9; Mismatches 70; Indels 42; Gaps 9;

OY 44 PGYHRLCCVSCIRKASGPNPGDPPTSLPFPSTGSPD-----PGQDADAAEPG 96
DB 128 PGR-----CITCPAGAGP-AGPPGAPGPKGNNGPAGAGAGSGGPGPPRGPADADASPG 182

OY 97 KPTGSEDHCHGRATQLGALDTSSTPGQHPFAPETPIPGASWSISPTTGGLPW-----G 151
DB 183 GP-----GHPGSPGNRGGQGRGT-----PGASGPPGPGAGAGGGRSG 226
OY 152 WTOTPTP--VPEDGPGGEDLHNGTSLPADLPGRPPEPCHPTTGLVLPDSDL 206
DB 227 GAGTTPGPGGPGSPGPG-----HSGND---GVETPTGNPAPGADAAVCPGARSSV 276

RESULT 6
ID 099LH3 PRELIMINARY; PRT; 724 AA.
AC 099LH3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIMILAR TO NADRIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC003259; AA003259.1; -;
DR HSSP: Q07960; IRGP.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR000198; RhogAP.
DR PRINTS: PR01217; PRICHXENSN.
DR SMART: SM00324; RhogAP.1.
SQ SEQUENCE 724 AA; 79434 MW; FF559ECED3A1CCAA CRC64;

Query Match 13.3%; Score 160; DB 11; Length 724;
Best Local Similarity 25.5%; Pred. No. 0.00012;
Matches 61; Conservative 19; Mismatches 95; Indels 64; Gaps 8;

OY 2 EGOWVPOSGPLHINKISTEPCTGDRSVFCOMEVLDRCYCSIRKASGP 61
DB 489 EGDLYVKESPPEKPKDSVAAPVAGRNS--NOMTTVPNOAGTGSNHSLSVTPHSAAGP 546
OY 62 NP-----GPPGPTSLPFPSTPGSPD-----PGQDADAAEPGKPTGSDHCH 106
DB 547 SPHTLRAVKKRPARPPKGNLPPGHPGQSSPGTSPKPSARSPSPQD--Q0000Q 604
OY 107 GRATQLGALDTS-----PGTQHPAPETPIPGASWSISPTTGGLP--- 149
DB 605 Q0000QPPGMRCSSSLPIQAPSHPPQPPRLGEGEPGPGTTPPTTPPTPLAK 664
OY 150 -----WGWOTPTTPVEDKGGQGEDLHNGTSLPADLPGRPPEPCHPTTGL 197
DB 665 QNPQSETTQLHGTLP RP RPVK-----PRNRBPVPPPP-HPGTHIV 706

RESULT 7
ID 09NX19 PRELIMINARY; PRT; 651 AA.
AC 09NX19;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE CDNA FLJ20219 FIS. CLONE COLF371.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON MUCOSA;
RA Matenabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Oobayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isonaga T., Sugano S.;

RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK000226; BAA91021.1; -
 DR InterPro: IPR002126; Cadherin.
 DR SMART: SM00112; CA; 1.
 DR PROSITE: PS50268; CADHERIN_2; 3.
 DR SEQUENCE 651 AA; 67379 MW; AE6D4984ED69175C CRC64;

Query Match 13.1%; Score 158; DB 4; Length 651;
 Best Local Similarity 26.1%; Pred. No. 0.00015;
 Matches 63; Conservative 5; Mismatches 69; Indels 104; Gaps 11;

QY 53 SCIRKASGPNPDPGPTSLPPFS-----TPG-----SPLPGQDPA 89
 DB 326 STTSSGEGTGHPPPGSTTLRPPTSSPGSGPAENSTSHQATPGDGTATKPKGTSPM 385
 QY 90 DAAPRGKPTGSEDDHGRATOLPGALDSSPGTQHPAP----- 129
 DB 386 ---PPGVGTST---SHQATPSGGTGTVPETGTSQMPMPMSGTSTSHQATPGGTAQT 438
 QY 130 -----ETPIP---GASWSISPTTPGGTLPWGTOTPTP-----VPEDKQPGEDL 170
 DB 439 PEAGTSQMPMPGKGTSTSHQPTTGG---GTAQTPEPGTSQMPPLSKTSPSSGGSPBDK 495
 QY 171 R-----HPGTSLPADLPGRPRECHPTGTFTLCVL 200
 DB 496 RFSVDMALGVLGALLLALLGLAVLVKHYGPRKLC-CGSKAPRP-OPQGFDMQAF 553
 QY 201 P 201
 DB 554 P 554

RESULT 8
 Q9HBB5 ID Q9HBB5 PRELIMINARY; PRT; 845 AA.

AC Q9HBB5: 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE MCDHL-FL.
 GN MCDHL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA "MCDHL, a Novel cDNA with Pacin and Cadherin-like Domains."
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF538676; AAG16733.1; -
 DR InterPro: IPR002126; Cadherin.
 DR SMART: SM00112; CA; 3.
 DR PROSITE: PS00232; CADHERIN_1; 1.
 DR PROSITE: PS50268; CADHERIN_2; 4.
 DR SEQUENCE 845 AA; 88156 MW; 10E654279CE30ECE CRC64;

Query Match 13.1%; Score 158; DB 4; Length 845;
 Best Local Similarity 26.1%; Pred. No. 0.00019;
 Matches 63; Conservative 5; Mismatches 69; Indels 104; Gaps 11;

QY 53 SCIRKASGPNPDPGPTSLPPFS-----TPG-----SPLPGQDPA 89
 DB 492 STTSSGEGTGHPPPGSTTLRPPTSSPGSGPAENSTSHQATPGDGTATKPKGTSPM 551
 QY 90 DAAPRGKPTGSEDDHGRATOLPGALDSSPGTQHPAP----- 129
 DB 552 ---PPGVGTST---SHQATPSGGTGTVPETGTSQMPMPMSGTSTSHQATPGGTAQT 604
 QY 130 -----ETPIP---GASWSISPTTPGGTLPWGTOTPTP-----VPEDKQPGEDL 170

DB 605 PEAGTSQMPMPGKGTSTSHQPTTGG---GTAQTPEPGTSQMPPLSKTSPSSGGSPBDK 661
 QY 171 R-----HPGTSLPADLPGRPRECHPTGTFTLCVL 200
 DB 662 RFSVDMALGVLGALLLALLGLAVLVKHYGPRKLC-CGSKAPRP-OPQGFDMQAF 719
 QY 201 P 201
 DB 720 P 720

RESULT 9
 Q9HAU3 ID Q9HAU3 PRELIMINARY; PRT; 845 AA.

AC Q9HAU3: 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE MU-PROTODCADHERIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-KIDNEY;
 RA Soileman A., Krieger S., Haase A., Hantusch B.;
 RT "Cloning and characterization of human mu-proteocadherin."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF301909; AAG33495.1; -
 DR InterPro: IPR002126; Cadherin.
 DR SMART: SM00112; CA; 3.
 DR PROSITE: PS00232; CADHERIN_1; 1.
 DR PROSITE: PS50268; CADHERIN_2; 4.
 DR SEQUENCE 845 AA; 88171 MW; D12C11C1E3E11680 CRC64;

Query Match 13.1%; Score 158; DB 4; Length 845;
 Best Local Similarity 26.1%; Pred. No. 0.00019;
 Matches 63; Conservative 5; Mismatches 69; Indels 104; Gaps 11;

QY 53 SCIRKASGPNPDPGPTSLPPFS-----TPG-----SPLPGQDPA 89
 DB 492 STTSSGEGTGHPPPGSTTLRPPTSSPGSGPAENSTSHQATPGDGTATKPKGTSPM 551
 QY 90 DAAPRGKPTGSEDDHGRATOLPGALDSSPGTQHPAP----- 129
 DB 552 ---PPGVGTST---SHQATPSGGTGTVPETGTSQMPMPMSGTSTSHQATPGGTAQT 604
 QY 130 -----ETPIP---GASWSISPTTPGGTLPWGTOTPTP-----VPEDKQPGEDL 170
 DB 605 PEAGTSQMPMPGKGTSTSHQPTTGG---GTAQTPEPGTSQMPPLSKTSPSSGGSPBDK 661
 QY 171 R-----HPGTSLPADLPGRPRECHPTGTFTLCVL 200
 DB 662 RFSVDMALGVLGALLLALLGLAVLVKHYGPRKLC-CGSKAPRP-OPQGFDMQAF 719
 QY 201 P 201
 DB 720 P 720

RESULT 10
 Q18751 ID Q18751 PRELIMINARY; PRT; 309 AA.

AC Q18751: 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE C50F7.5 PROTEIN.
 GN C50F7.5
 OS Caenorhabditis elegans.

		13.0%;	Score 156.5;	DB 5,	Length 309;
	Best Local Similarity	34.1%;	Pred. No. 9.6e-05;		
Matches	56;	Conservative	10;	Mismatches	55;
				Indels	43;
				Gaps	11.
QY	59	SGNPGPDGPTSLPFFSTP---	GSLPLGGDDPADAAEPCKPKTGSSEDHGNRATQLPG	115	
Db	102	SGP---PSGPNV--PSEDPPQSGPRSPGVDSDDQSVSR--	SDDHPS-GPSPGG	153	
QY	116	LIDS-----SCTGTCHPAPEPTIPGASWSISITPTGGLFWCMTQTTP	156		
Db	154	VDSSEDPQSVSESDPPQSGPSPGPVPD--SSDDPPSSSSRGVPDSDEP--	SPSGPP	210	
QY	157	TYPEDKQGQGEDLR-----HPGSTLPAD--LGRPRECHNG	193		
Db	211	SPGPVD--PSDDPKRSEPSPGVPDESDESPSDPPGPAPPG	251		

DR PROSITE; PS50268; CADHERIN_2; 4.
SQ SEQUENCE 845 AA; 88097 MW; 6CD0E3985391F59F CRC64;

	Query Match	Score	DB	Length
	Best Local Similarity	26.1%	Pred. No.	0.00029;
Matches	63; Conservative	4; Mismatches	70; Indels	104; Gaps
QY	53 SCIKKASGPNPGDDEGPISLPPFS-----TPG-----SPLYGGDDPA	89		
Db	492 STTSSSGGTGRPHRPSGCTTLRRPTSTSGPSPAENKSTSHOAPATGGDTAOTPKRGTSOPM	551		
QY	90 DAAREPGKPTGSEDHONGFATOLPALTISSPGTONHAPAR-----	129		
Db	552 -----PRGVGTST--SHOAPATSSGGTAOTPRRPGTSQPMPRPBSMGSTSHOAPATGGCGTAOT	604		
QY	130 -----ETPLP---GASWSISPTTPGCLRWMTOTPTP-----VPEDKGAGCEDL	170		
Db	605 PEAGTSQPMPPGMGTGSTSHOPTTBGC---GTAQTPEFGTSQPMWLSKSTSSGGGPSIEDK	661		
QY	171 R-----HPGSLPADLPGRPREPCPTPTFTLCVL	200		
Db	662 RFESVDMALGSVIGALLLLALLGLAVLVNHHYGRFLKC-CSGKAPEP-QPGSFGDMAQL	719		
QY	201 P 201			
Db	720 P 720			

	RESULT	12	
ID	Q09N39		
AC	Q09N39	PRELIMINARY;	PRT; 768 AA.
DT	01-JUN-2001	(TRENBLrel. 17, Created)	
DT	01-JUN-2001	(TRENBLrel. 17, Last sequence update)	
DT	01-DEC-2001	(TRENBLrel. 19, Last annotation update)	
DE	NADRINZ.		
GN	NMADRINZ.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus		
OX	NCBI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=BRAIN;		
RA	Kobayashi Y., Furuta B., Itoh K., Umeda M.;		
RT	"Mus musculus homolog of rat Nadrin."		
RL	Submitted (APR-2001) to the EMBL/GenBank/DDBB databases.		
DR	EMBL: AB060554; BAB43663.1; -		
DR	HSSP: Q07960; IRGP.		
DR	InterPro: IPR002965; P_Rich_extensn.		
DR	InterPro: IPR000198; RhOGAP.		
DR	Pfam: PF006620; RhOGAP. 1.		
DR	SMART: PR01217; PRICHEXTENSNM.		
DR	SMART: SM00324; RhOGAP. 1.		
DR	SEQUENCE 768 AA; 84309 MW; C1CF1IC73CA7044F CRC64;		

[illegible]

DR InterPro: IPR002486; COL_cuticle_N.
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF01484; COL_cuticle_N; 1.
 SO SEQUENCE 305 AA; 30586 MW; F24770C129694718 CRC64;

Query Match 12.6%; Score 151.5; DB 5; Length 305;
 Best Local Similarity 31.3%; Pred. No. 0.00023;
 Matches 71; Conservative 12; Mismatches 59; Indels 85; Gaps 17;

QY 7 PQSGPLPINKISTEPTGDRSVFCOMENVLDRCSTPGYHRLCCVCSICKKASGPNCGPD 66
 DB 71 PRSNPF-----QSLYRQKRLPDVC-----ICQPLEINAPRP-RGP- 106
 67 PGPSLP-----PFS--TPGSPLPGPD-----PADAAEPGK--PTGSEDHQ 106
 107 PGPPGQPGHPEQPGHVGQPGSGPAPPCPLPQQAQRCRPAAGAGTGTGKQGPAG-QPGP 165
 107 GRATQPLGALDTSS-----PGTHPFAPETPT-----PGASWSISPTTGGGLPMG 151
 166 GR-----PGAPKSSGAGPAGPAGPAGPAGKNGRPGQPGKNGVSHPTTIPG--PKG 219
 QY 152 WTQPTPTVPEDKGGPGEDELRRHPTSLPADLPGR-----PRECHPTG 193
 DB 220 PGSGP-----GQPK--PG--PAGVAGKTGPGPGPQVPGAG 252

RESULT 16

Q9WRM2 PRELIMINARY; PRT; 448 AA.

AC Q9WRM2; 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE LATENT NUCLEAR ANTIGEN.
 OS Macaca mulatta rhadinovirus 17577.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 NCBI_TaxID=83534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99174001; PubMed=10074154;
 RA Searles R.P., Berguam E.P., Axthelm M.K., Wong S.W.;
 RT Sequence and genomic analysis of a rhesus macaque rhadinovirus with
 RT similarity to Kaposi's sarcoma-associated Herpesvirus/Human
 RT herpesvirus 8.";
 J. Virol. 73:3040-3053(1999).
 [2]
 SEQUENCE FROM N.A.
 Searles R.P., Berguam E.P., Axthelm M.K., Wong S.W.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083501; AAD21406.1;
 DR InterPro: IPR002965; P_rich_extensn.
 DR InterPro: IPR000531; TonB_boxC.
 DR PRINTS: PRO1217; PRICHEXTENS.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
 SO SEQUENCE 448 AA; 47459 MW; E69BE03FEE4A974F CRC64;

Query Match 12.5%; Score 151; DB 12; Length 448;
 Best Local Similarity 27.6%; Pred. No. 0.00036;
 Matches 63; Conservative 10; Mismatches 71; Indels 84; Gaps 15;

QY 1 PEGOWPQSGPLP-----INKISTEPTGDR-SVFCOMENVLDRCSTPGYHRLCCVCSCT 55
 DB 132 PASHSP 173
 QY 56 KKASGPNCGPDGPTSLPPTSTPGSP-----IPGQPDADAAEPGKPTGSEDHQ 108
 DB 174 ---DSBP--PEP--PTSLPPTDSGPPSPPTSPSPSPSPSPSPSPSPSPSPSPSPSP 222
 QY 109 ATQLPGLADLSSPGTQ-----HPAPF---TPRG-----ASWSISPTTGGGLP- 149

DB 223 -----APSPNQAVSHTDPTGSPKPCPPPGHTSHSYVGWG--PPTRAGVPC 272
 QY 150 -----MGWTPPTPPE-DKQPGEDLRHPTSLPADLGRPREPC 191
 DB 273 LRLRGTSHSHDEDAPERQDEGERQD-----FARPPRPPR 312

RESULT 17

Q93207 PRELIMINARY; PRT; 316 AA.

AC Q93207; 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE C15A11.6 PROTEIN.
 GN C15A11.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gardner A.E.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z79694; CAB01958.1;
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR002486; COL_cuticle_N.
 DR Pfam: PF01391; Collagen; 3.
 DR Pfam: PF01484; COL_cuticle_N; 1.
 SO SEQUENCE 316 AA; 31232 MW; 1FC26A3770E9A6D5 CRC64;

Query Match 12.5%; Score 150.5; DB 5; Length 316;
 Best Local Similarity 30.9%; Pred. No. 0.00028;
 Matches 56; Conservative 13; Mismatches 73; Indels 39; Gaps 12;

QY 44 PGYN-RLCVCSTCKKASGPNP--DPGPTSLPPTSTPGSP-----PGQDDPADAAEP 94
 DB 125 PSHHQKQECIKCPQGLPGAGAPGQPGPG--PNCNPGAPAGGGGQPGPGPGAGS 182
 QY 95 PGKP-----TGSEDHQHGATQLPGALDTSSPTGHPFAPETP--IPGASWSISPTTGG 148
 DB 183 PGAGAPGPNPSPGSGQGRGLPG--PSGAGSGP--PGAPGPGGGMNAPGAPPG- 237
 QY 149 PNGWTQPTPTVPEDKGGPGEDELRRHPTSLPADLPGRPREPCHTPTFTLCVLPDLSQLRG 208
 DB 238 PAG-----PNCQPG--HPGQD--GQCAPGNDGTPGSDAAYCPCPTRSSVLR 280
 QY 209 H 209
 DB 281 H 281

RESULT 18

Q19763 PRELIMINARY; PRT; 301 AA.

AC Q19763; 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE F23H12.4 PROTEIN.
 GN F23H12.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Kershaw J.K.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "genome sequence of the nematode C.elegans: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z74472; CA98942.1; -
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR002486; COL_cuticle_N.
 DR Pfam: PF01484; COL_cuticle_N; 1.
 DR PIR: P01484; COL_cuticle_N; 1.
 SQ SEQUENCE 301 AA; 29177 MW; 4A75B65540E50360 CRC64;

Query Match 12.4%; Score 149.5; DB 5; Length 301;
 Best Local Similarity 34.9%; Pred. No. 0.00032;
 Matches 59; Conservative 6; Mismatches 69; Indels 35; Gaps 13;

QY 51 CVCSTKASGPNPDP-----DPGPTSLP-----PFTSGSPLP-GPDDPADAEPPKPTG 100
 DB 147 CKPCPGPGRPPR-PGRPARPDGAGRPGRGTDAAAGSGRPGPAGAGAPG-PAG 204
 QY 101 SEDHGKATQLPGALDTSSTPGQHPPAPETP-IPGASMSISPTTGGLPWGTQTPTTY 159
 DB 205 E-----TPAISEPLTPGAPGEPGDSGPPGPPGAP-GNDGPPGP- 246
 QY 160 PEDKGQGEDLRHPTSLRPLDGRPERCHPTGTFILCVLPDSOLRG 208
 DB 247 PGPAGAPGP-GPPVDGSGSPRG-PGPAGTGEKGC-PRYCALDG 291

RESULT 19
 Q9Y2M4 PRELIMINARY; PRT; 995 AA.

AC Q9Y2M4;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ATROPHIN-RELATED PROTEIN ARP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20195630; PubMed=10729226;
 RA Anker L.C., Bauer A., Corvi R., Dhlmann S., Praml C., Cavenee W.K.,
 Schwab M., Hampton G.M.;
 RT "Identification and characterization of novel genes located at the
 t(1;15)(p36.2;q24) translocation breakpoint in the neuroblastoma cell
 line NGP."
 RL Genomics 64:195-202(2000).
 DR EMBL: AF118275; AAD27584.1; -
 DR InterPro: IPR002951; Atrophin.
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PRO1222; ATROPHIN.
 DR PRINTS: PRO1217; PRICHEXTENSIN.
 SQ SEQUENCE 995 AA; 107088 MW; 888B827A44BFB04 CRC64;

Query Match 12.4%; Score 149.5; DB 4; Length 995;
 Best Local Similarity 31.2%; Pred. No. 0.001;
 Matches 50; Conservative 9; Mismatches 74; Indels 27; Gaps 6;

QY 64 GPDGPTSLPFTSGSPRLPGDPAADAEPPKPTGSEDHQHGKATQLPGALDTSSTPGT 123
 DB 175 GVPAPSSAPP-GTPQLTPGTPPSATVPPQGSPTASQ-----APNQPAPTAPVPH 227

QY 124 Q-----HPFAPETPIPGASWSISP-----TPGGLPWGWTQTPTTPVEDKQPGED-- 169
 DB 228 HIQAPALHPQRPSPHPPHPPHPLPLQPLTGSAGQPSASHPQL-HQGGPPGHSL 286
 QY 170 -----LRHPTSLPADLGRPPPECHPTGTYTLCVLPDS 204
 DB 287 QAGPLQHHPGPPOFGLPPQASQOAPLGTSPAAAYPHTS 326

RESULT 20

Q43393 PRELIMINARY; PRT; 1012 AA.
 AC Q43393;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE ATROPHIN-1 RELATED PROTEIN.
 GN DRPLA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Xia J.-H., Liu C.-Y., Ruan Q.-G., Wang D.-A., Deng H.-X.;
 RT "Cloning and localization of Human Atrophin-1 (DRPLA) related gene."
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF016005; AAC31120.1; -
 SQ SEQUENCE 1012 AA; 108967 MW; 4DMA8872A6B78866 CRC64;

Query Match 12.4%; Score 149.5; DB 4; Length 1012;
 Best Local Similarity 31.2%; Pred. No. 0.001;
 Matches 50; Conservative 9; Mismatches 74; Indels 27; Gaps 6;

QY 64 GPDGPTSLPFTSGSPRLPGDPAADAEPPKPTGSEDHQHGKATQLPGALDTSSTPGT 123
 DB 192 GVPAPSSAPP-GTPQLTPGTPPSATVPPQGSPTASQ-----APNQPAPTAPVPH 244
 QY 124 Q-----HPFAPETPIPGASWSISP-----TPGGLPWGWTQTPTTPVEDKQPGED-- 169
 DB 245 HIQAPALHPQRPSPHPPHPPHPLPLQPLTGSAGQPSASHPQL-HQGGPPGHSL 303
 QY 170 -----LRHPTSLPADLGRPPPECHPTGTYTLCVLPDS 204
 DB 304 QAGPLQHHPGPPOFGLPPQASQOAPLGTSPAAAYPHTS 343

Search completed: August 29, 2002, 00:11:10
 Job time: 421 sec

Thu Aug 29 10:46:29 2002

us-09-938-330-22_copy_1040_1249.rpt



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 19:56:43 ; Search time 33.31 Seconds
(without alignments)
918.068 Million cell updates/sec

Title: US-09-938-330-20
Perfect score: 6954
Sequence: 1 MAPRLALSLPLHLCALCT.....PTGTFLLVLPDSQRLRHT 1252

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgnt2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgnt2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgnt2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgnt2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgnt2_6/ptodata/2/1aa/5A_COMB.pep:*
6: /cgnt2_6/ptodata/2/1aa/5B_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296.5	4.3	855	4	US-09-813-819-2
2	296.5	4.3	855	4	US-09-920-048-2
3	294.5	4.2	814	4	US-09-813-819-4
4	294.5	4.2	814	4	US-09-920-048-4
5	291	4.2	441	3	US-08-985-526-3
6	272	3.9	696	4	US-09-351-414-2
7	261	3.8	529	2	US-08-836-442-3
8	261	3.8	751	4	US-08-836-443-3
9	256	3.7	464	4	US-09-411-329C-14
10	247	3.6	462	4	US-09-411-329C-3
11	247	3.6	462	4	US-09-411-329C-17
12	242	3.5	735	2	US-08-765-243-6
13	242	3.5	735	2	US-08-765-243-6
14	238.5	3.4	769	1	US-08-243-542-4
15	238.5	3.4	769	1	US-08-477-407-4
16	238.5	3.4	769	1	US-08-484-355-4
17	235.5	3.4	1170	1	US-08-313-288B-20
18	234	3.4	734	2	US-08-765-243-8
19	234	3.4	734	2	PCT-US95-07295-8
20	231	3.3	651	1	US-08-264-101-2
21	231	3.3	651	1	US-08-765-243-2
22	231	3.3	651	1	PCT-US95-07295-2
23	229.5	3.3	391	4	US-08-706-216-6
24	215	3.1	788	2	US-08-918-914-4
25	214	3.1	470	3	US-08-813-150-2
26	213	3.1	470	3	US-08-514-014-2
27	213	3.1	470	2	US-08-833-823-2

28	212.5	3.1	681	2	US-08-655-345-4	Sequence 4, Appl1
29	212.5	3.1	681	3	US-09-183-275-4	Sequence 4, Appl1
30	212.5	3.1	681	5	PCT-US96-08407-4	Sequence 4, Appl1
31	212.5	3.1	807	2	US-08-655-345-2	Sequence 2, Appl1
32	212.5	3.1	807	2	US-09-183-275-2	Sequence 2, Appl1
33	212.5	3.1	807	5	PCT-US96-08407-2	Sequence 2, Appl1
34	210.5	3.0	670	1	US-08-243-542-3	Sequence 3, Appl1
35	210.5	3.0	670	1	US-08-477-407-3	Sequence 3, Appl1
36	210.5	3.0	670	1	US-08-484-355-3	Sequence 3, Appl1
37	206.5	3.0	1172	1	US-08-313-288B-19	Sequence 3, Appl1
38	205	2.9	469	1	US-08-313-288B-15	Sequence 19, Appl1
39	202.5	2.9	239	5	PCT-US93-01652-1	Sequence 15, Appl1
40	198	2.8	748	4	US-09-030-335-9	Sequence 1, Appl1
41	195	2.8	749	2	US-08-937-931-8	Sequence 8, Appl1
42	195	2.8	749	4	US-08-285-502-8	Sequence 8, Appl1
43	195	2.8	749	4	US-09-126-8	Sequence 8, Appl1
44	195	2.8	807	1	US-07-862-021B-10	Sequence 10, Appl1
45	195	2.8	807	1	US-08-313-288B-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1
US-09-813-819-2
Sequence 2, Application US/09813819
Patent No. 6294368
GENERAL INFORMATION:
APPLICANT: MERCKUOY, Gennady et al
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: C0001177
CURRENT APPLICATION NUMBER: US/09/813,819
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 855
TYPE: PRT
ORGANISM: Human
US-09-813-819-2

Query Match 4.38; Score 296.5; DB 4; Length 855;
Best Local Similarity 23.7%; Pred. No. 1.1e-13;
Matches 185; Conservative 72; Mismatches 288; Indels 237; Gaps 43;
QY 53 LSHVSGPAAASAGSMVDTPTLPRHSHLRVARSPLHPG-----GLMP 98
|||
Db 35 LSWGLPAGG-----TEEOAESEKAPRPLEPVOYLODDPLSLKKVLOTSLP 83
QY 99 GRVGRSHLYRNIVFEGKEHLRLRPNRRLVPGSSVEW--QEDFRELPQPLROBCVYTG 156
|||
Db 84 -----EPLRIKLELDGDSHLELLONRRELVPGRPILWYQPDGTRVSGHTLENCYGG 138
QY 157 GVTGMPAAVAIAISNCGIAGL-IRISTDFIE-----PIERGOEKASARTNAV 206
|||
Db 139 RVRKYAGSWSICTGSLKRLVLLPERSSTLYEGGCDLGGPITISRIQDLHLPGHTAL 198
QY 207 YRREAV--QOEAMPEPDGLINEAFGLDPLNLGLVGDLDGTEKKRRHAKGVSIEVL 264
|||
Db 199 SWRESVHTQRPPEHPGGRH-----IKRRDVTETKTVELV 235
QY 265 LAYVDS-VVAFHKEKVQNVYTLTANIYDEIYHDSLGVINIALVRLIMGYRQSLSI 323
|||
Db 236 IVADHSEAKYRDFQHLNRLTLEVALLDTFPP-----LNVFVALVGEAWTOR---DLV 288
QY 324 E-RGNRSRLIEQYCW--AHSQRODPSSHAHHNDHVVLETRDF-GPS-GYAPYTGKCHP 378
|||
Db 289 EISPNVAITLENLTHRRKHLPLP-----HDSAQVLTGTSFGPTVGMAIONISICSP 342
QY 379 LRSCLNHEDEGS---SAFVIAHETGHVLMGMEHDGCGACDETSLGSMVPLVOAFAHR 435

```

Db 343 DFGGVNMDHSTSLIGVASSIAHELGHSLGLDHLDPGNSC-----PCGPA--- 388
QY 436 FHMRSCKLELSRYLPSYDCLDDPPDPAMPPELPGINYS-----MDQCFDFGSGY 490
Db 389 -----PAKTCIMEASTD-----FLPGLNFSNCSRALEKALD---GM 423
QY 491 QTCIAFRT-----FEPCKQLWCSHPDNPYFCKTKKGPPLDGTGC--APGKW 534
Db 424 GSCLEFRLPSLPRMAAFGCMNFVEPEGDCDGLD---C---VDPCCDSLTCOLRPAQ 477
QY 535 CFK-GHCIMKSPQTYGQDGGSWTKFGSCS--RSCGGVRSRSCNNPSPAYG-GRP 590
Db 478 CASDGPCC---QNCOLRPSGWOCRPTRKDDCLPEFCBGD---SSQC--PPDVSIGDEP 528
QY 591 CLGPMFEYOVNSSECPGTEDFRAOOC-----AKRNSY-YVHQN 629
Db 529 CAGGQ---AVCMHGRG-ASY---AQCCOSLMGPGAQPAAPLCLOTANTRGNAFSGCGRN 580
QY 630 AKHSWVYEPDD-----DAQKCELICQ--SADTGDVY 659
Db 581 PGGSYVSCTPRDAICGQLCQGTGRTQPLLSIRDLMTETIDVNGTELNCMSVHLDLGSVY 640
QY 660 FMMQVYHDGTRGSTRDPYVCARGECVPYGCDEKVEGSMKADKC---GYCGDNSHCRTV 716
Db 641 AQLPLTLPGTAGC---PGLVCIDHRCQRYDL---LGAQCRSKCHGHVC-DSNRHCYCE 693
QY 717 KG 718
Db 694 EG 695

```

```

RESULT 2
US-09-920-048-2
; Sequence 2, Application US/09920048
; Patent No. 6344352
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001177DIY
; CURRENT APPLICATION NUMBER: US/09/920,048
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/813,819
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Human
; 09-920-048-2

```

```

Query Match 4.3%; Score 296.5; DB 4; Length 855;
Best Local Similarity 23.7%; Pred. No. 1,1e-13;
Matches 185; Conservative 72; Mismatches 288; Indels 237; Gaps 43;

QY 53 LSHVYSGPAAASAGSMVYDPTPLPRHSHSLRVASPLHPG-----GTLWP 98
Db 35 LFMGVLGPAGG-----TEEQAESEKAPREPPEVYLODLPISLKKVLOTSLP 83
QY 99 GAVGRISLYFNTVYGKEHLRLRNRLVYVGSSEVEM--QEDFRELROPLROCVYTG 156
Db 84 -----EPLRIKLELDGDSHILELONRELVEGPRPLVWYQDPGTVYSEGHLEMCCTOG 138
QY 157 GVTMGCAVAVAISNCDGLAGL-IRTDSTDFIE-----PLERGOEKESGRTYHV 206
Db 139 RVAGYAGSVNSICTGSGGLAGLVLLPERSYTLEGGPGLDQPSLIRIQLDLPLPHTCAL 198
QY 207 YRREAV--QOEWAEPDGDILHNEAFGLGDLPLNLGLVGDLGSTEKKRRRAKGSISIVL 264

```

```

Db 199 SMRESVHTQTPPEHPLQGRH-----IRRRRDVYETETKVELV 235
QY 265 LVYDDS-VYRPHGKEHYOVNYLTLMNIYDELYIHDESLGHNINILVYLIMVYGRQSLI 323
Db 236 IYADHSEAKYKRDQHLNRLTEVALLDITFEP---LNVRLVGLGEATQR---DLV 288
QY 324 E-RGNPSRSLBOYCRW--AHSQORODPSHAEDHVVFLTRPQD-GPS-GYAPYGMCHP 378
Db 289 EISNPAYLTLENFLMKRAHLLPLRP-----HNSAQLVYTSFSGPYGMAIONSICSP 342
QY 379 LRSCALNHEDGFS--SAFVIAHETGHVLMGHEHOGNGCADEFTSLGSVAPLYOAAFHR 435
Db 343 DFGGVNMDHSTSLIGVASSIAHELGHSLGLDHLDPGNSC-----PCGPA--- 388
QY 436 FHMRSCKLELSRYLPSYDCLDDPPDPAMPPELPGINYS-----MDQCFDFGSGY 490
Db 389 -----PAKTCIMEASTD-----FLPGLNFSNCSRALEKALD---GM 423
QY 491 QTCIAFRT-----FEPCKQLWCSHPDNPYFCKTKKGPPLDGTGC--APGKW 534
Db 424 GSCLEFRLPSLPRMAAFGCMNFVEPEGDCDGLD---C---VDPCCDSLTCOLRPAQ 477
QY 535 CFK-GHCIMKSPQTYGQDGGSWTKFGSCS--RSCGGVRSRSCNNPSPAYG-GRP 590
Db 478 CASDGPCC---QNCOLRPSGWOCRPTRKDDCLPEFCBGD---SSQC--PPDVSIGDEP 528
QY 591 CLGPMFEYOVNSSECPGTEDFRAOOC-----AKRNSY-YVHQN 629
Db 529 CAGGQ---AVCMHGRG-ASY---AQCCOSLMGPGAQPAAPLCLOTANTRGNAFSGCGRN 580
QY 630 AKHSWVYEPDD-----DAQKCELICQ--SADTGDVY 659
Db 581 PGGSYVSCTPRDAICGQLCQGTGRTQPLLSIRDLMTETIDVNGTELNCMSVHLDLGSVY 640
QY 660 FMMQVYHDGTRGSTRDPYVCARGECVPYGCDEKVEGSMKADKC---GYCGDNSHCRTV 716
Db 641 AQLPLTLPGTAGC---PGLVCIDHRCQRYDL---LGAQCRSKCHGHVC-DSNRHCYCE 693
QY 717 KG 718
Db 694 EG 695

```

```

RESULT 3
US-09-813-819-4
; Sequence 4, Application US/09813819
; Patent No. 6294368
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001177
; CURRENT APPLICATION NUMBER: US/09/813,819
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 814
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(814)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-813-819-4

```

```

Query Match 4.2%; Score 294.5; DB 4; Length 814;
Best Local Similarity 24.2%; Pred. No. 1,4e-13;
Matches 170; Conservative 67; Mismatches 258; Indels 207; Gaps 40;

QY 119 LRLRNRLVYVGSSEVEM--QEDFRELROPLRQECVYTGVTGMGPAVAVAISNCDGLAG 176

```

```

Db      82 LELLQNRRLVPRPPLVWYQDPGRVYSEGTLENCYQGRVRYAGSWSICTGSLRG 141
QY      177 L-IRTDSTDFIE-----PLERGOEKEASGRTHVYRRAV--QOEMAPDGLH 224
Db      142 LVVLTPERSYSTLEQPGDLOGPPIISRIDLHGHICALSWRESVHTQRPPEHPLGQRH 201
QY      225 NEAFGLGDLPNLILGVGDQLDGTERRKRRHAKPGSYSTEVLVYVDS-VYFHKKEHYQNT 283
Db      202 -----IRRRDVTETKVELYELVADHSEAKYRDFQHLNR 238
QY      284 VLTLMNIVDEITYHDESLGVHINIALVRLIMVGRSLSTIE-RGNPSRSLQYCRW-AH 340
Db      239 TLEVALLDITFEFRP-----LVNRYVALVGLAMTOR--DIVEISPNPAVLLENFLHWRRAH 291
QY      341 SQQRDPDSHAHHHDHVVFLTRQDF-GPS-GYAPVTGCMCHLRSCALNHEDGFS--SAFY 395
Db      292 LILRLP-----HDSADLVGTSTSPGPTVMATONISICSPDSSGVNMDHSTSLGVASS 345
QY      396 IAHETGVLGMEHDGNGCAGDETSLSGVMAPIVQAFFHFRHWSRCSKLELSRYLPSTYDC 455
Db      346 IAHETGSLGIDHDLPGNSC-----PCGPA-----PAKTC 376
QY      456 LLDPPRPAMPQPELPGIYNS-----MDQCRPDESGYOTCLAFRT-----498
Db      377 TMEASTD-----FLPGLNFSNCSRRALKEALD--GMSCLFRLRLSLPMAAFQCN 426
QY      499 --FEPCKOLMCSHDNPFCKTKGPRLDGTEC--APGKMCFK-GHICIMKSPQTYGGDG 553
Db      427 MFVBERGQCCGFLD---C---VDPCCDSLTCQLRGAQASDGPCC---ONQLRPS 476
QY      554 GWSWTKEFGSCS--RSCGGGVRSRSCNNPSPAYG-GRPCLGPMFEYOVCNSEECPGTY 610
Db      477 GWCORPTRGCDLPERCPGD-----SSQC--PRDVSIDGEPGACGQ--AVCHMGRG-ASY 527
QY      611 EDFAAQC-----AKRNSY-VYHONAKHSWYPERPD-----641
Db      528 ----AQCQSLMGPGAOPAPRLCLQTANTRGNMFGSCGRNPSGSYVCTPRDALCGQLQC 583
QY      642 -----DAQKCELICQ--SADTGDVFMNQVYVHDGTRCSYRDPYSY 679
Db      584 QTRGTQPLIGSIRDILMETIDVNGTELNCWSVHLDLGSVQAPLILTPGTACG--PGLV 640
QY      680 CARCECPVPGCKEYSGMKADKDC--GYCGDNSHCRTYKG 718
Db      641 CIDHRCQAVDL---LGAQECRSKCHGHGVC-DSNRHCYCEEG 678
BOLT 4
S-09-920-048-4
; Sequence 4, Application US/09920048
; Patent No. 6344352
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001177DIY
; CURRENT APPLICATION NUMBER: US/09/920,048
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/813,819
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 814
; TYPE: PRT
; ORGANISM: Human
; FEATUERE:
; NAME/KEY: VARIANT
; LOCATION: (1)....(814)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-920-048-4

```

```

Query Match      4.2%: Score 294.5; DB 4; Length 814;
Best Local Similarity 24.2%; Pred. No. 1,4e-13;
Matches 170; Conservative 67; Mismatches 258; Indels 207; Gaps 40;

QY      119 LRLPRNRLVWPGSSVEM--QEDFRELFPRLROECVYTGGVGMGAVALISNCDGLAG 176
Db      82 LELLQNRRLVPRPPLVWYQDPGRVYSEGTLENCYQGRVRYAGSWSICTGSLRG 141
QY      177 L-IRTDSTDFIE-----PLERGOEKEASGRTHVYRRAV--QOEMAPDGLH 224
Db      142 LVVLTPERSYSTLEQPGDLOGPPIISRIDLHGHICALSWRESVHTQRPPEHPLGQRH 201
QY      225 NEAFGLGDLPNLILGVGDQLDGTERRKRRHAKPGSYSTEVLVYVDS-VYFHKKEHYQNT 283
Db      202 -----IRRRDVTETKVELYELVADHSEAKYRDFQHLNR 238
QY      284 VLTLMNIVDEITYHDESLGVHINIALVRLIMVGRSLSTIE-RGNPSRSLQYCRW-AH 340
Db      239 TLEVALLDITFEFRP-----LVNRYVALVGLAMTOR--DIVEISPNPAVLLENFLHWRRAH 291
QY      341 SQQRDPDSHAHHHDHVVFLTRQDF-GPS-GYAPVTGCMCHLRSCALNHEDGFS--SAFY 395
Db      292 LILRLP-----HDSADLVGTSTSPGPTVMATONISICSPDSSGVNMDHSTSLGVASS 345
QY      396 IAHETGVLGMEHDGNGCAGDETSLSGVMAPIVQAFFHFRHWSRCSKLELSRYLPSTYDC 455
Db      346 IAHETGSLGIDHDLPGNSC-----PCGPA-----PAKTC 376
QY      456 LLDPPRPAMPQPELPGIYNS-----MDQCRPDESGYOTCLAFRT-----498
Db      377 TMEASTD-----FLPGLNFSNCSRRALKEALD--GMSCLFRLRLSLPMAAFQCN 426
QY      499 --FEPCKOLMCSHDNPFCKTKGPRLDGTEC--APGKMCFK-GHICIMKSPQTYGGDG 553
Db      427 MFVBERGQCCGFLD---C---VDPCCDSLTCQLRGAQASDGPCC---ONQLRPS 476
QY      554 GWSWTKEFGSCS--RSCGGGVRSRSCNNPSPAYG-GRPCLGPMFEYOVCNSEECPGTY 610
Db      477 GWCORPTRGCDLPERCPGD-----SSQC--PRDVSIDGEPGACGQ--AVCHMGRG-ASY 527
QY      611 EDFAAQC-----AKRNSY-VYHONAKHSWYPERPD-----641
Db      528 ----AQCQSLMGPGAOPAPRLCLQTANTRGNMFGSCGRNPSGSYVCTPRDALCGQLQC 583
QY      642 -----DAQKCELICQ--SADTGDVFMNQVYVHDGTRCSYRDPYSY 679
Db      584 QTRGTQPLIGSIRDILMETIDVNGTELNCWSVHLDLGSVQAPLILTPGTACG--PGLV 640
QY      680 CARCECPVPGCKEYSGMKADKDC--GYCGDNSHCRTYKG 718
Db      641 CIDHRCQAVDL---LGAQECRSKCHGHGVC-DSNRHCYCEEG 678
RESULT 5
US-08-985-526-3
; Sequence 3, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Connolly, Boyle, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
08-985-526-3

```

```

Query Match          4.2%; Score 291; DB 3; Length 441;
Best Local Similarity 20.1%; Pred. No. 1,1e-13;
Matches 119; Conservative 49; Mismatches 176; Indels 248; Gaps 24;

```

```

QY 473 GINSMDEQ-----CRFDGSGYOTCLAFRTFPCQOLMCSHPDNFYCKTKKPLDGT 527
DB 22 GVQYRNNEWTDVSC-----TECHCQNSVYICKKVC--PIMP--CSNATVP--DG- 66
QY 528 ECAPGKWCFFGHCIWSP----- 545
DB 67 ECCPRCPMSADADMGSPSEMTSCSTSCGNIQGRSCDSLNRCGESSVOTRCHIQ 126
QY 546 --EQTGODGWSWTKFSGSCSGGVSRSRSCNNPSPAYGAPCLGPMFEYOVCS 603
DB 127 ECDRFKODGWSHWSWSSCSTYCGGVITRITLCSPPQNGKCEBARETACKK 186
QY 604 EECGTYEDFPAQOCARNSYVYHONAKHSWVPEPD-----DAQCELIQOSA-- 653
DB 187 DACP-----INGGWGPMSPMDICSVTCGGGVQKRSRLCYDSRM 224
QY 654 DTGVVFNQV-----VHOSTRCSYRDPYSVYCARGECVPGCDKEVSMKADKGVCG 707
DB 225 TEENKELANELRRPPLCYHNGVYRNNEWTV-----DSCTEC- 262
QY 708 GDNSHCRTVKGTLGKASKQAGALXVQIPAGARHIQIEALEKSPHSVKNQVTFILN 767
DB 263 --HCQNSVYICKKVCSPIMPSCNATVPDG----- 289
QY 768 PKGKATISRTITAMGLEMEADAVEAKESLTSGLPFLAIALPTEGGRSSLAYKV 827
DB 290 --ECCPRCPMS-----DSADG----- 304
QY 828 IHEDLPLIGSNVLEEMDTYEMA-LKSNAPCSKACGGGIGFTKYGC--RRRRDHNV 883
DB 305 -----WSPMSEMTSCSTSCGNGIQGRSCDLSLNRCGESSV 341
QY 884 QRHLCDHKRRPKPIRRCONHPSCQPVV--TEEWAGCSRGSGKIGVOTRIGQCLPLSLNG 942
DB 342 QTRCHIQE-----CDKRFKODGWSHWSWSSCSTVTCGD--GVYTR-----ITLCNS 387
QY 943 THKVPAPACAGDRPEARPCRLVPCPQWHLGAMSQ--CSATCGEGIGOR 991
DB 388 PSPQNGKPCBEGARET-KACKKACPLNGGNGPSPMDICSVTCGGGVQKR 438

```

```

RESULT 6
US-09-351-414-2
; Sequence 2, Application US/09351414
; Patent No. 6265199

```

```

; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baindur, Nand
; APPLICANT: Delsner, Theresa A.
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: DISINTEGRIN HOMOLOG
; FILE REFERENCE: 98-29
; CURRENT APPLICATION NUMBER: US/09/351,414
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-351-414-2

```

```

Query Match          3.9%; Score 272; DB 4; Length 696;
Best Local Similarity 21.3%; Pred. No. 5.3e-12;
Matches 149; Conservative 102; Mismatches 247; Indels 200; Gaps 35;

```

```

QY 108 FNVYFGKELHLRLPNRLVPGSSVWEQDEFLRQPLR-----QECVYTGAVTG 160
DB 30 FOIEAFGSKFILLDLINNGLL-----SSDYVEIHENKRPQYSGKEHCYHGSIRG 81
QY 161 MGAVALAISNDGLAGLIRTDSTDFEFLERGOOEKASGRTHVYVRYREAVQOEMAPP 220
DB 82 VADSVYALSTGCKLHGMEFDDTFYMIPELPLYNDEK-STGRPHLIQKTLAGYS----- 135
QY 221 GDLINEAFGLDPLNLGLVGDQ-----LDDTERKRNHAP--GSYS-----EVLVVD 268
DB 136 -----KOMKMLTMRGDQWPFLELQMLKRRKAVNPSSKIGIEEMKYLELMGND 185
QY 269 DSVAFPH--GKEHYQNYVLTLMNIVDELYHDESLGVHINIALVRLIMVYRQSLIERG 326
DB 186 HRTYKHKHRSNAHTNNFAKSVNLVDSTYKEQ-----LNTRVLVAVETWEXQIDITT 240
QY 327 NPSRSLQVYCRMAHSSQOROPSHAEHNDHVVFLTRQDF--GPSGYAPYTGACHPRLSCAL 384
DB 241 NPVQMLHEFSKY--RQR-----IKOHADAVHLISHVTFHYKRSLSLYEGVCSRTRGVY 293
QY 385 NHEDGFSSAF--VIAHETGHVLGMEHDGQ-----NGC-ADETSIGSVAPLVQ 430
DB 294 N-EYGLPMVAQVLSQSLAQNLGLOMEPSSRKPKCDCTESMGCLMETGVS----- 344
QY 431 AAFHREHMSRCSKLELSRYLP--SYDCLLDDP--FDPAMPQPELPQINYSMDEQCRFD 485
DB 345 --HSRKFSGKSLIEYRDFLQRGCGACLENRPYKLFEPY-----EC--- 382
QY 486 FSGYOTCLAFRTFPCQOLMCSHPDNFYFCKTKKGRPLDTEGAPGKWCFFGHCIW--- 542
DB 383 -GNGY-----VEAGEBCDCGFHYECYGLCKCKCSLSNGAHGSDGPPCCNNTSCLPDR 433
QY 543 -----KSPQYTGODGWSWTKFSGCSRSG--GGVRSRS 576
DB 434 GYICRAVAVNECDITTEYCTGDSGCCPPNHLKQD-----YACNOGRCRYNGCKTRD 485
QY 577 RSCNNPSPAYGAPCLGPMFEYOVCSSECPGTIEDFPAQOCARNSYVYHONAKH--- 632
DB 486 NQCY--IWGTAKAGSDKFCYEKLNTF--GTEKQ--NCGKAGDRTI-QCSKHDVFC 534
QY 633 -----SWVPEPDQKQKELI-----COSA-----DTGDVYFNQVYVHDTGR 670
DB 535 GFLLCNTLFRAPRIGOLGELIPTSTYHOGRYVYIDSGAHVAVLDDTDIVY--VEDGTP 590
QY 671 CSYRDPYVACARGEVAVGCDKEVSGMKADKGVCGG 708
DB 591 CG---PSMCLDRKCLQIOA-LMSSCPLDRKGVCSG 624

```

```

RESULT 7
US-08-836-442-3

```


Query Match	3.8%	Score 261	DB 2	length 529;
Best Local Similarity	23.3%	Pred. No. 2.4e-11		
Matches 132;	Conservative 78;	Mismatches 262;	Indels 94;	Gaps 28;
QY	79	HSSTLRVARSPLPGGITLMKPRGVKSHSLFNVATYFGKELHLRLPNRRRLVPG-SSYEWQ	137	
Db	37	HLHLKRRKAKPEEQ-----PQETELTKYKKTITNGKLAIVLYLKNNLLAAGTTEYYNA	89	
QY	138	EDRELPEKPD-LRQECVYTGVTGMPGAVALSNQDLAGLIRDSDFEIEPLERGOOE	196	
Db	90	STGEITTSPOIMDDCYOGHILINKEKVSADSIKRGISGQDGRYFIEPLSPITHR	149	
QY	197	KEASGRTHVYVYRREAAVOQEMAEPPDDLNEAFSGLC-----DLNLLGLVDOLGDTERR	250	
Db	150	----GQEHAFK-----YNPDKRYNDSCTGMGVLMALDLOONALPATKL--VKLR	195	
QY	251	RRHAKGSYSIEVLIVDDSVVRFH--GKEHYQNVLTLMNITVEIYIHDSLSGVHINIAL	308	

```

RESULT      8
US-08-836-443-3
: Sequence 3, Application US/08836443
: Patent No. 5883241
:
GENERAL INFORMATION:
APPLICANT: DOCHERTY, Andrew, J.P
APPLICANT: SLOCOMBE, Patrick, M.
TITLE OF INVENTION: DNA SEQUENCES CODING FOR A
TITLE OF INVENTION: HUMAN METALLOPROTEINASE AND VARIANTS THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BROWSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,443
FILING DATE: 01-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9709420
FILING DATE: 05-SEP-1997
APPLICATION NUMBER: 9612145.4
FILING DATE: 11-JUN-1996
APPLICATION NUMBER: 9526230.9
FILING DATE: 21-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47424
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 751 amino acids
TYPE: amino acid
STRADEDNESS: single

```

TOPOLOGY: linear
MOLECULE TYPE: NO. 5883241e
US-08-836-443-3

Query Match 3.8%; Score 261; DB 2; Length 751;
Best Local Similarity 25.1%; Pred. No. 3.8e-11;
Matches 136; Conservative 67; Mismatches 201; Indels 136; Gaps 36;

247 TERRRRHAKPGSY-----IEVLLVDDSVYRF-----HGKEHVONYVLTLMNIVDEIYHDE 298
19 TKKPRRRKRRDLSNMKYVELLYVAD---YLEFQKNRRDQATGKHKLIEIANYDKFYR-- 74
299 SLGVHINAIYRLIMVGYROSLSLIERGNSRSLIYQYCRMAHSQOQODPSHAHHHVVYR 358
75 --SINIRAIYGLVWTHGNMCVSE--NRYSTL---WFLSWRRKLLAQYKHDAQL 125
359 LTRDQGPS--GYAPVTGMCPLRSCALN---HEDGFSAFVIAHETGHVIGMEHDSQGN 413
126 ITGMSFHGTITGLAPLAKMSYVSGGVNMDHSENAIGVATMAHEMGNFQWTHDSADC 185
414 GCADETSIGSYMAPLVQAARFHH--WSRCSKLELSRYLPSYD--CLDDPDPAMPQPP 469
186 CSASADGCGTMA---AATGHPFKVFNCGNRRELDRLYQSGGCMCLSNMP-DTRMLYGG 241
470 ELPGINTSMD-EQCRDFGSGYQ---TCLAFR-TEEP---CKQIMCSHP----- 510
242 RRCGNGTIEDEEC--DGEERECNPPCCNANCTLRPGAECAGSCCHQCKLLAPGLIC 299
511 -----DNPFYCKTKKGP-----PLDTECAPGK-WCFKGHCITKSPQYIGDQGW 555
300 REQARQCDLPFC-TGKSPHCPITNFYQMDGTPCEGQAYCYNQCL-----TYQE--- 348
556 SSWTFGSGSRSCTGGVARSRSRSC-----NNSPAYG--GRCCIGPMNEYOVYCNSECP-G 608
349 -----QCOQIMGFGARPADLCEKYNVAGDTFGNCGKMMN---EHRCKNRDCKG 398
609 TYEDFRAQCCAKRNSYHYHONAKSHWYVEPDDADOKCELICQSA-----DTGDIVVFM 661
399 KI-----QCCSEARPLESNA---VPIDITILMNGQIQCRGTHYV RGP EEGDMLDP 448
662 NOVYHDTGRCTRDYPSYVANGEC-----VPVCGDKFVSGMKADKDC--GVCGGP-NS 711
449 G-LVMTGTRKCGYN--HICFEGQCRNMTSFFETEGCGK-----KCNHGVCNMMQNC 495
712 HC 713
496 HC 497

RESULT 9
US-09-411-329C-14
Sequence 14; Application US/09411329C
Patent No. 6261820
GENERAL INFORMATION:

APPLICANT: Boone, Thomas
APPLICANT: Li, Huimin
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
FILE REFERENCE: A-596
CURRENT APPLICATION NUMBER: US/09/411.329C
CURRENT FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14

LENGTH: 464
TYPE: PRT
ORGANISM: Agkistrodon contortrix
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Native pro-fibrinolase of Agkistrodon contortrix
US-09-411-329C-14

Query Match 3.7%; Score 256; DB 4; Length 464;
Best Local Similarity 24.7%; Pred. No. 4.6e-11;
Matches 98; Conservative 67; Mismatches 158; Indels 74; Gaps 19;

92 PGCTLMPGRVGRHSILYFVNYVFGKELHLRLRPNRLVYVG--SSVEMQEDFRELRLQ 150
114 PRGAVQPKY--EDAMQYEFKVNSEPVVLLHKKKGLFSEDSYETHSPDGREITTYPLGE 171
151 E-CVYTGVTGMPGAVALISNCDGLAIRDSIDFIEPLERGOQOEAGSRTHVYRR 209
172 DHCYHGRLENDADASTASACNGLKGFKLQGEWYLIPELSDSE-----AHAVYKY 225
210 EAVQOEWAEPDGLHNEAFGLDLPNLLIGVD-----QLGPTERRRRHAKPGS 258
226 ENVEKE-----DEAPKMGVGNMSEYPIKAPQNLTKRQRF--PQR 268
259 YSIEVLLVDDSV-VREHG-KEHVONYVLTLMNIVDEIYHDESGLVHINAIYRLIMGY 316
269 Y-VQVIYVADHMMNTRKYNQSDSKIRQWVQIYVNTINEIYRP---LNTQFTLVGLEISN 323
317 ROSLSLIERGNSRSLIYQYCRMAHSQOQODPSHAHHHVVYFLTRDQGPS--GYAPVTG 374
324 QDLITVTSYSHD--TLASFGNM---RETDLLRQRHDAQLTIAIDPDGTVGLAYVG 377
375 MCHPLRSCALNHEDGFSS-----AFVIAHETGHVIGMEHDSQNGCADETSIGSYM 425
378 MCQ-----LKHSTGVIOHSAINLLVALLTMAHELGHNHMDGNOCGANS---CVM 428
426 APYVQAARFHFMSRCSKLELSRYLP--SYDCLDDP 460
429 AAMLSDQPSKL-FSDCSKDYQTFITVNNPCLINKP 464

RESULT 10
US-09-411-329C-3
Sequence 3; Application US/09411329C
Patent No. 6261820
GENERAL INFORMATION:
APPLICANT: Boone, Thomas
APPLICANT: Li, Huimin
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
FILE REFERENCE: A-596
CURRENT APPLICATION NUMBER: US/09/411.329C
CURRENT FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 462
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Native pro-NAT (analog of fibrinolase)
US-09-411-329C-3

Query Match 3.6%; Score 247; DB 4; Length 462;
Best Local Similarity 24.4%; Pred. No. 2.1e-10;
Matches 97; Conservative 66; Mismatches 158; Indels 76; Gaps 19;

92 PGCTLMPGRVGRHSILYFVNYVFGKELHLRLRPNRLVYVG--SSVEMQEDFRELRLQ 150
114 PRGAVQPKY--EDAMQYEFKVNSEPVVLLHKKKGLFSEDSYETHSPDGREITTYPLGE 171
151 E-CVYTGVTGMPGAVALISNCDGLAIRDSIDFIEPLERGOQOEAGSRTHVYRR 209
172 DHCYHGRLENDADASTASACNGLKGFKLQGEWYLIPELSDSE-----AHAVYKY 225
210 EAVQOEWAEPDGLHNEAFGLDLPNLLIGVD-----QLGPTERRRRHAKPGS 258
226 ENVEKE-----DEAPKMGVGNMSEYPIKAPQNLTKR---SFPQR 266

```

QY 259 YSIEVLVVDSV-VREHG-KEHVOYVLTLMNIVDEIYHDESLGVHINIALVRLIMGY 316
Db 267 Y-VOLVIVADHRMNTKYNQSDSDKIRQWVHIVNTINELIRP---LNIQFTVGLIEMSN 321
QY 317 RQSLIERGNPSRSLQVQCRMAHQOQROPSHAHHDHVFLTRQDFGPS--GYAPVTG 374
Db 322 QDLITVTSVSHD--TLASFGNW---RETDLRQRQHRDHNAQLTALIDFDGDTVGLAYVG 375
QY 375 MCHPLRSCALNHEDGESS-----AFVIAHETGHVIGMEHHDGQNGCADETSLSGYM 425
Db 376 MQQ-----LKHSTGVITODHSAINLVALTWAEHLGHNHGDGNOCHGANS---CYM 426
QY 426 APVQAFHFRHWSRCSKTELRYLP--SYDCLDDP 460
Db 427 AAMLSDPSKL-FSDCSKKDYQYTFIVNPNQCLINKP 462

```

Search completed: August 28, 2002, 20:11:08
 Job time: 865 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 28, 2002, 23:53:34 ; Search time 32.43 Seconds

(without alignments)
158.168 Million cell updates/sec

Title: US-09-938-330-22_COPY_1040_1249

Perfect score: 1206
Sequence: 1 PEGQWPGSGPLHPIINKISS.....PTGPTLCVLPDLSQRLGHT 210

Indexing table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /usr2/6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCRTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175	14.5	761	2	US-08-707-237A-84
2	175	14.5	762	1	US-08-642-255-114
3	175	14.5	762	1	US-08-642-255-120
4	175	14.5	762	1	US-08-397-633A-26
5	175	14.5	762	1	US-08-397-633A-31
6	158	13.1	682	1	US-08-642-255-126
7	158	13.1	682	1	US-08-397-633A-36
8	154	12.8	417	1	US-08-175-155-69
9	154	12.8	417	1	US-08-477-509B-104
10	154	12.8	417	1	US-08-642-255-102
11	154	12.8	417	2	US-08-707-237A-76
12	154	12.8	417	3	US-08-482-085B-104
13	154	12.8	417	4	US-08-444-791A-104
14	154	12.8	837	1	US-08-175-155-68
15	154	12.8	837	1	US-08-642-255-103
16	154	12.8	837	1	US-08-642-255-101
17	154	12.8	837	2	US-08-707-237A-75
18	154	12.8	837	2	US-08-482-085B-103
19	154	12.8	837	4	US-09-444-791A-103
20	154	12.8	837	4	US-08-397-633A-50
21	152.5	12.6	829	1	US-08-642-255-132
22	152.5	12.6	829	1	US-08-397-633A-53
23	151.5	12.6	960	4	US-09-219-849-5
24	150	12.4	330	1	US-08-642-255-32
25	150	12.4	408	1	US-07-609-716-65
26	150	12.4	408	4	US-08-475-411A-65
27	150	12.4	408	4	US-08-478-029A-65

28	145.5	12.1	1255	2	US-09-080-897-4	Sequence 4, Appl
29	145.5	12.1	1255	3	US-08-899-595-1	Sequence 1, Appl
30	145.5	12.1	1255	4	US-09-323-735-4	Sequence 4, Appl
31	145	12.0	623	4	US-09-029-348-3	Sequence 3, Appl
32	145	12.0	626	4	US-09-029-348-2	Sequence 2, Appl
33	143.5	11.9	357	1	US-07-609-716-66	Sequence 66, Appl
34	143.5	11.9	357	1	US-08-642-255-33	Sequence 33, Appl
35	143.5	11.9	357	4	US-08-475-411A-66	Sequence 66, Appl
36	143.5	11.9	357	4	US-08-478-029A-66	Sequence 66, Appl
37	137	11.4	1248	2	US-09-080-897-2	Sequence 2, Appl
38	137	11.4	1248	4	US-09-323-735-2	Sequence 2, Appl
39	137	11.4	1341	3	US-08-963-825-18	Sequence 18, Appl
40	137	11.4	1341	4	US-09-570-573-18	Sequence 18, Appl
41	137	11.4	1341	4	US-09-548-608-18	Sequence 18, Appl
42	136	11.3	1442	2	US-08-316-650-12	Sequence 12, Appl
43	136	11.3	1442	5	PCT-US95-02251-12	Sequence 12, Appl
44	135	11.2	1274	4	US-09-095-443-2	Sequence 2, Appl
45	134.5	11.2	1315	3	US-08-899-595-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-707-237A-84
Sequence 84, Application US/08707237A

Patent No. 5830713

GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A.

APPLICANT: Capello, Joseph

APPLICANT: Crisman, John W.

APPLICANT: Dorman, Mary A.

TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC

NUMBER OF SEQUENCES: 108

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/707,237A

FILING DATE: 03-SEP-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/175,155

FILING DATE: 29-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/053,049

FILING DATE: 22-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/609,716

FILING DATE: 06-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/269,429

FILING DATE: 09-NOV-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/114,618

FILING DATE: 29-OCT-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 06/927,258

FILING DATE: 04-NOV-1986

ATTORNEY/AGENT INFORMATION:

NAME: Treacartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-55186-10/MHD

Db 87 GOOHHLGARGQ---AGDVGSPG-----APGTPGQGLPGSPG--APGTPGQGLPGSPGA 136
 QY 153 TOTPTP-----VPEKGGPGEGLRH-----PGTSLPADLPGRPPEPCHP 191
 Db 137 PGTTPGQGLPGSPGAPGTTPGEGQOHHLGARGAGDVGSPGAPGTTPGQGLPGSPGAPGTP 196

RESULT 4
 US-08-397-633A-26
 ; Sequence 26, Application US/08397633A
 ; Patent No. 5773577
 ; GENERAL INFORMATION:

APPLICANT: Cappello, Joseph
 TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATE-ESCAPABLE
 TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
 NUMBER OF SEQUENCES: 105
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/397, 633A
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Rowland, Bertlam I
 REGISTRATION NUMBER: 20,015
 REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELETYPE: 910 277299
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 762 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 8-397-633A-26

Query Match 14.5%; Score 175; DB 1; Length 762;
 Best Local Similarity 35.0%; Pred. No. 2.2e-06;
 Matches 63; Conservative 8; Mismatches 51; Indels 58; Gaps 12;

QY 58 ASGP-----NPGPD-----PGTSLPFPSTPGSP--LPGQDPAADAEPPKPTG 100
 Db 29 ASDPMGAPGTTPGQGLPGSPGAPGTTPGQGLP--GSPGAPGTTPGQGLPGSPGAPGTTPG 86
 QY 101 SEDHNGRATQRLGALDTSSPGTQHPFAPETP----IPGASWSISPTTPG--GLPW--GW 152
 Db 87 GOOHHLGARGQ---AGDVGSPG-----APGTPGQGLPGSPG--APGTPGQGLPGSPGA 136
 QY 153 TOTPTP-----VPEKGGPGEGLRH-----PGTSLPADLPGRPPEPCHP 191
 Db 137 PGTTPGQGLPGSPGAPGTTPGEGQOHHLGARGAGDVGSPGAPGTTPGQGLPGSPGAPGTP 196

RESULT 5
 US-08-397-633A-31
 ; Sequence 31, Application US/08397633A
 ; Patent No. 5773577
 ; GENERAL INFORMATION:
 APPLICANT: Cappello, Joseph

TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATE-ESCAPABLE
 TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
 NUMBER OF SEQUENCES: 105
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/397, 633A
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Rowland, Bertlam I
 REGISTRATION NUMBER: 20,015
 REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELETYPE: 910 277299
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 762 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-397-633A-31

Query Match 14.5%; Score 175; DB 1; Length 762;
 Best Local Similarity 35.0%; Pred. No. 2.2e-06;
 Matches 63; Conservative 8; Mismatches 51; Indels 58; Gaps 12;

QY 58 ASGP-----NPGPD-----PGTSLPFPSTPGSP--LPGQDPAADAEPPKPTG 100
 Db 29 ASDPMGAPGTTPGQGLPGSPGAPGTTPGQGLP--GSPGAPGTTPGQGLPGSPGAPGTTPG 86
 QY 101 SEDHNGRATQRLGALDTSSPGTQHPFAPETP----IPGASWSISPTTPG--GLPW--GW 152
 Db 87 GOOHHLGARGQ---AGDVGSPG-----APGTPGQGLPGSPG--APGTPGQGLPGSPGA 136
 QY 153 TOTPTP-----VPEKGGPGEGLRH-----PGTSLPADLPGRPPEPCHP 191
 Db 137 PGTTPGQGLPGSPGAPGTTPGEGQOHHLGARGAGDVGSPGAPGTTPGQGLPGSPGAPGTP 196

RESULT 6
 US-08-642-255-126
 ; Sequence 126, Application US/08642255
 ; Patent No. 5773249
 ; GENERAL INFORMATION:
 APPLICANT: Cappello, Joseph
 TITLE OF INVENTION: HIGH MOLECULAR WEIGHT COLLAGEN-LIKE
 TITLE OF INVENTION: Protein Polymers
 NUMBER OF SEQUENCES: 135
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 126:
SEQUENCE CHARACTERISTICS:
LENGTH: 682 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
8-642-255-126

Query Match 13.1%; Score 158; DB 1; Length 682;
Best Local Similarity 37.6%; Pred. No. 4.3e-05;
Matches 62; Conservative 9; Mismatches 50; Indels 44; Gaps 14;

QY 58 ASGP-----NPGPD-----PGPTSLPPTSTPGSP-LPGOPDPAADAEPPGKPTG 100
DB 29 ASDPMGAPGTPOGGLPGSPGAPGTTPGGLP--GSPGAGCTGPGGLPSPGAPGTPGG 86
QY 101 SEDHGHGRATQLPGLDITSSPTGQ---HPAPETP---IPGASWSISPTTG--GLPW 150
DB 87 AK--QAGDVGS--PGAPGT--PGPGLPGSPGAPGTTPGGLPGSPG--APGTTPGGLPG 139
QY 151 --GWTQTPTPVPE--DKGPGEDLRHPTSLPADLPGRPPEPCHP 191
DB 140 SPGAPGTGAGAKQAGDVGSPGA---PGTTPGGLPGSPGAPGTP 180

RESULT 7
US-08-397-633A-36
Sequence 36, Application US/08397633A
Patent No. 5773577
GENERAL INFORMATION:
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE
TITLE OF INVENTION: OF ENZYMAIC CROSS-LINKING
NUMBER OF SEQUENCES: 105
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/397,633A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A-58648-1/BIR PROP-011-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 682 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-397-633A-36

Query Match 13.1%; Score 158; DB 1; Length 682;
Best Local Similarity 37.6%; Pred. No. 4.3e-05;
Matches 62; Conservative 9; Mismatches 50; Indels 44; Gaps 14;

QY 58 ASGP-----NPGPD-----PGPTSLPPTSTPGSP-LPGOPDPAADAEPPGKPTG 100
DB 29 ASDPMGAPGTPOGGLPGSPGAPGTTPGGLP--GSPGAGCTGPGGLPSPGAPGTPGG 86
QY 101 SEDHGHGRATQLPGLDITSSPTGQ---HPAPETP---IPGASWSISPTTG--GLPW 150
DB 87 AK--QAGDVGS--PGAPGT--PGPGLPGSPGAPGTTPGGLPGSPG--APGTTPGGLPG 139
QY 151 --GWTQTPTPVPE--DKGPGEDLRHPTSLPADLPGRPPEPCHP 191
DB 140 SPGAPGTGAGAKQAGDVGSPGA---PGTTPGGLPGSPGAPGTP 180

RESULT 8
US-08-175-155-69
Sequence 69, Application US/08175155
Patent No. 5641648
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
APPLICANT: Crissman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: Methods for Preparing Synthetic
TITLE OF INVENTION: Repetitive DNA
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-175-155-69


```

Query Match          12.88; Score 154; DB 1; Length 837;
Best Local Similarity 35.98; Pred. No. 0.00011;
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;
XY-----NPPPD-----PQPTLPPFTPGSP-LPGQPDADAEPGKRP- 98

```

DB 29 ASDPMGAPGTGPGQLPGSPGAPGTGPGQLP--GSPGAPGTGPGQLPGSPGAPGTGPG 86
QY 99 ---TSEEDHOGHGRATQLPALDTSSTPGTQ---HPFAPETP-----IPGASMSISPTTGP- 146
DB 87 QGLPGS-----PGAPGT--PGPQGLPGSPGAPGTGPGQLPGSPG--APGTGPG 131
QY 147 -GLPMGWTQTPPYVED-KGQGEDLRHGTSLPADLPGRPPEPCHP 191
DB 132 QGLPGSPGAPGTGPGQLPGSPGA-----PGTGPQGLPGSPGAPGTG 174

RESULT 15

US-08-477-509B-103
; Sequence 103, Application US/08477509B
; Patent No. 5770697

GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A
APPLICANT: Cappello, Joseph
APPLICANT: Crisman, John W
APPLICANT: Dorman, Mary A
TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,509B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-509B-103

Query Match 12.8%; Score 154; DB 1; Length 837;
Best Local Similarity 35.9%; Pred. No. 0.00011;
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;

QY 58 ASGP-----NPGPD-----PGPTSLPPTSTPGSP-LPGQDPPADAAEPCKP-- 98
DB 29 ASDPMGAPGTGPGQLPGSPGAPGTGPGQLP--GSPGAPGTGPGQLPGSPGAPGTGPG 86
QY 99 ---TSEEDHOGHGRATQLPALDTSSTPGTQ---HPFAPETP-----IPGASMSISPTTGP- 146
DB 87 QGLPGS-----PGAPGT--PGPQGLPGSPGAPGTGPGQLPGSPG--APGTGPG 131
QY 147 -GLPMGWTQTPPYVED-KGQGEDLRHGTSLPADLPGRPPEPCHP 191
DB 132 QGLPGSPGAPGTGPGQLPGSPGA-----PGTGPQGLPGSPGAPGTG 174

RESULT 16

US-08-642-255-101
; Sequence 101, Application US/08642255
; Patent No. 5773249

GENERAL INFORMATION:

APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Berttram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-101

Query Match 12.8%; Score 154; DB 1; Length 837;
Best Local Similarity 35.9%; Pred. No. 0.00011;
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;

QY 58 ASGP-----NPGPD-----PGPTSLPPTSTPGSP-LPGQDPPADAAEPCKP-- 98
DB 29 ASDPMGAPGTGPGQLPGSPGAPGTGPGQLP--GSPGAPGTGPGQLPGSPGAPGTGPG 86
QY 99 ---TSEEDHOGHGRATQLPALDTSSTPGTQ---HPFAPETP-----IPGASMSISPTTGP- 146
DB 87 QGLPGS-----PGAPGT--PGPQGLPGSPGAPGTGPGQLPGSPG--APGTGPG 131
QY 147 -GLPMGWTQTPPYVED-KGQGEDLRHGTSLPADLPGRPPEPCHP 191
DB 132 QGLPGSPGAPGTGPGQLPGSPGA-----PGTGPQGLPGSPGAPGTG 174

RESULT 17
US-08-707-237A-75
; Sequence 75, Application US/08707237A
; Patent No. 5830713
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Capello, Joseph
; APPLICANT: Crissman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
; TITLE OF INVENTION: REPETITIVE DNA
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fleht, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/707,237A
; APPLICATION NUMBER: US/08/707,237A
; FILING DATE: 03-SEP-1996
; CLASSIFICATION: 435
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-10/WHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-707-237A-75

Query Match 12.8%; Score 154; DB 2; Length 837;
Best Local Similarity 35.9%; Pred. No. 0.00011;
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;

OY 58 ASGP-----NRGP-----PGPSTLPSTPGSP-LPGPDADAAEPCKP-- 98
DB 29 ASDPMGACGTGGPGLPGSPGAPGTGGPGLP--GSPGACGTGGPGLPGSPGAPGTGP 86

OY 99 ---TGSEHDHGRATGALDITSSPTQ-----HPPAETP-----IPGASMSISPTPG- 146
DB 87 QGLPGS-----GGAGT--PGPGLPGSPGAPGTGGPGLPGSPG--ABGTGCP 131

OY 147 -GLPWGWTGTTPVPED-KGQGEDLRHPTSLPADLPGRPPRCPH 191
DB 132 QGLPGSPGACGTGGPGLPGSPGAPGTGGPGLPGSPGAPGTGP 174

RESULT 18
US-08-482-085B-103
; Sequence 103, Application US/08482085B
; Patent No. 6018030
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Richardson, Charles
; APPLICANT: Chambers, James
; APPLICANT: Causey, Stuart
; APPLICANT: Pollack, Thomas J.
; APPLICANT: Crissman, John W.
; TITLE OF INVENTION: No. 6018030e1 Peptides Comprising Repetitive
; TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
; NUMBER OF SEQUENCES: 112
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fleht, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,085B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-482-085B-103

Query Match 12.8%; Score 154; DB 3; Length 837;
Best Local Similarity 35.9%; Pred. No. 0.00011;
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;

Thu Aug 29 10:46:18 2002

us-09-938-330-22_copy_1040_1249.ra

Page 11

Db 132 0G1PGSPGAPGTPGPGQGLPGSPGA---PGTPPGQGLPGSPGAPGTP 174

Search completed: August 29, 2002, 00:08:25
Job time: 891 sec

alignment_scores:	Length:
Quality: 6723.50	1220
Ratio: 5.525	Gaps:
Percent Similarity: 89.43%	Percent Identity: 99.26%

alignment_block:
 US-09-938-350-20 x AF358666 ..

2000
14 w/s inserted at

2 2000 fished a
+ 17 was washed a

[illegible]

```

467 nProGluLeuProGlyIleasnTyrSerMetAspGluGlnCysArgp 484
|||||
1401 GCCCCAGAGCTCCTGGATGACACTCAATGATGACAGTGCCT 1450
484 heaSpheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheGlu 500
|||||
1451 TTACTTTGGCAGTGGCTACAGACCTGCTTGCCATTCCAGACCTTTAG 1500
501 ProCysLysGlnLeuThrPysSerHisProAspAsnProTyrPheCysLy 517
|||||
1501 CCTGGACAGCAGCTGTGTGACAGCCATCTGCACACCCGTACTTCTGCAA 1550
517 sThrLysGlyProProLeuAspGlyThrGluCysAlaProGlyLysT 534
|||||
1551 GACCAAGAGAGGGCCCCCGCTGGATGGAGCTGATGTCACCCGGCAAGT 1600
534 rPcysPheLysGlyHisCysIleTrrPlysSerProGluGlnThrTyrGly 550
|||||
1601 GGTGCTTCAAGGTCATGTCATCTGGAAGTCGCCGAGCAGACATATGCG 1650
551 GlnAspGlyGlyTrrPserSerTrpThrLysPheGlySerCysSerArgSe 567
|||||
1651 CAGGATGAGAGCTGAGCTCCCTGGACCAAGTTGGGTATGTTCGGGTC 1700
567 rCysGlyGlyGlyAlaArgSerArgSerArgSerCysAsnAsnProSerP 584
|||||
1701 ATGTGGGGCGGGGTGCATCCCGCAGCGGAGGTGCACACACCCCTCC 1750
584 roAlaTyrGlyGlyArgProCysLeuGlyProMetPheGluTyrGlnAla 600
|||||
1751 CAGCCTATGAGAGCGCCCGCTGTAGGGCCCATGTCAGTACCAAGTGC 1800
601 CysAsnSerGluGluCysProGlyThrTyrGlnAspPheArgAlaGlnG 617
|||||
1801 TCACAACAGAGAGAGTGCCTCGGACCTACAGACACTTCCGGGCCAGCA 1850
617 nCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSerT 634
|||||
1851 GTGTGCCAAGGCAACTGCTACTATGTGCACCAAGATGCCAAGCACAGCT 1900
634 rPValProTyrGluProAspAspAlaGlnLysCysGluLeuIleCys 650
|||||
1901 GGGTCCCTACGAGCTGACGATACGCGCCAGAAGTGTGAGCTGATCTGC 1950
651 GlnSerAlaAspTrpGlyAspValValPheMetAsnGlnValValHisAs 667
|||||
1951 CAGTCGGCGGACACGGGGACGTGTTCATGAAACCAAGTGGTTCACGA 2000
667 pGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArgLys 684
|||||
2001 TGGACACGCTGCACTACCGGAGCCATACAGCGTCTGTGCCGTGGCG 2050
684 lCysValProValGlyCysAspLysGluValGlySerMetLysAlaAsp 700
|||||
2051 AGTGTGTGCTGTGCGCTGTGACAAAGAGGTGGGTCCATGAAGGGCGAT 2100
701 AspLysCysGlyValCysGlyGlyAspAsnSerHisCysArgThrValLy 717
|||||
2101 GACAAAGTGTGAGTCTGCGGGGTGACAACTCCACTGCGAGGACTGTGAA 2150
717 sGlyThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuValG 734
|||||
2151 GGGAGCGCTGGGCAAGGCTCCCAAGCAGGAGAGAGCTCTCAACCTGTGC 2200
734 lInleProAlaGlyAlaArgHisIleGlnIleGlnAlaLeuGlnLysSer 750
|||||
2201 AGATCCAGAGGTGCCAGGCACATCCAGATTAGGCACATGGAGAGTCC 2250
751 ProHisArgSerValValLysAsnGlnValThrGlySerPheIleLeuAs 767
|||||
2251 CCCACCGCATTTGTGTGAAGAACAGGTACCGGCACTTCATCTCTCAA 2300
767 nProLysGlyLysGluAlaThrSerArgThrPheThrAlaMetGlyLeuG 784
|||||
|||||
2301 CCCCAAGGGCAAGGAGCCACAAAGCGGACTTCAACCCCATGGGCTCG 2350
784 lInrPcLysAspAlaValGluAspAlaLysGlnSerLeuLysThrSerLy 800
|||||
2351 AGTGGAGAGATGGGTGGAGATGCGCAAGAAAGCTCCAAAGCACAGCGGG 2400
801 ProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGlnGlyG 817
|||||
2401 CCCCTGCTGAAGCCATTGCCATCTCGGCTCTCCCGCCCACTGAGGGTGG 2450
817 yProAlaGSerSerLeuAlaTyrLysTyrValIleHisGluAspLeuLeuP 834
|||||
2451 CCCCCGACACAGCTGGCTTACAAATGATCATCATATGAGACCTGTGTC 2500
834 roLeuIleGlySerAsnAsnValLeuLeuGlnGluMetAspThrTyrGlu 850
|||||
2501 CCTTATCGGGAGCAACAATGCTCTCTGAGAGAGATGACACCTATGAG 2550
851 TrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGlyGlyI 867
|||||
2551 TGGGGCTTCAAAGAGCTGGGCCCTGCAGCAAGGCTGTGGAGAGGAT 2600
867 eGlnPheThrLysTyrGlyCysArgArgArgArgAspHisHisMetValG 884
|||||
2601 CCAAGTTCACCAATATGCGCTCCGGCGCAGACGAGACCAACCATGTGTC 2650
884 lInArgHisLeuCysAspHisLysLysArgProLysProIleArgArgArg 900
|||||
2651 AGCGACACCTGTGTGACCAACAAGAAAGAGCCCAAGCCATCCGCGGCGC 2700
901 CysAsnGlnHisProCysSerGlnProValTrpValThrGlnGlnTrpGl 917
|||||
2701 TCACAACAGCACCCGCTGCTCACTGTGTGAGTGCAGGAGAGTGGGG 2750
917 yAlaCysSerArgSerCysGlyLysLeuGlyValGlnThrArgGlyIleG 934
|||||
2751 TGCCTCGACCGGAGACTGTGTGGAGCTGGGGGTGCACACAGGGGGATAC 2800
934 lnCysLeuLeuProLeuSerAsnGlyThrHisLysValMetProAlaLys 950
|||||
2801 AGTGCTGCTGCTCCCTCTCCAAATGAAACCCCAAGATCAAGCGGCGCAA 2850
951 AlaCysAlaGlyAspArgProGluAlaArgArgProCysLeuArgValPr 967
|||||
2851 GCCTGTGCGGGGACCGGCTGAGAGCGCGGACCGCTGTCTCCGAGTGGC 2900
967 oCysProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAlaThrC 984
|||||
2901 CTGCCAGACCCAGTGGAGGCTGGGAGGCTGTGCCAGTGTCTGCCACCT 2950
984 ysglyGlyGlyIleGlnGlnArgGlnValAlaCysArgThrAsnAlaAsn 1000
|||||
2951 GTGAGAGAGGAGATCCACAGCGGAGAGTGTGTGCAGGACCAAGCGCAAC 3000
1001 SerLeuGlnHisCysGluGlyAspArgProAspThrValGlnValCysSe 1017
|||||
3001 AGCCTCGGGCATTTGCGAGGGGATAGGCCAGACACTGTCCAGGTCTGCAG 3050
1017 rLeuProAlaCysGlyGlyAsnHisGlnAsnSerThrValArgAlaAspV 1034
|||||
3051 CCTGCTGCTGCTGTGGAGAAATCACCAAGATCCAGGAGGAGGCGGATG 3100
1034 aLrrPcLysGlyLysThrProGluGlyGlyTrpValProLysSerGlyPro 1050
|||||
3101 TCTGGGACACTTGGAGCGCCAGAGGGGAGTGGGTGCCAACAATCTAACCC 3150
1051 LeuHisProIleAsnLysLysSerSerMetCysAlaAlaGlnProCysTh 1067
|||||
3151 CTACATCCCATTAACAAGATATCATCA.....ACGAGACCTGTGCAC 3191
1067 rGlyAspArgSerValPheCysGlnMetGluValLeuAspArgTyrCys 1084
|||||

```

3192 GGGAGACAGCTCTGCTTTCGCCAGATGGAAGTGTGATGCTACTGCT 3241
 1084 erileprogllytrhiargleucysvalsercysileylsalsala 1100
 3242 CCAATCCCGGCTACACCGGCTCTGCTGTCTGCTGATCAGAAAGGCC 3291
 1101 SerGlyProAsnProGlyProAspProGlyProThrSerLeuProProph 1117
 3292 TCGGGCCCCAACCTCGGCCAGACCTGGGCCCAACCTCACTGCCCTT 3341
 1117 eserThrProGlySerProleuProGlyProGlnAspProAlaAspAla 1134
 3342 CTCACGCTCCGGAAGCCCTTACAGAGACCCAGACCTGCGATGCTG 3391
 1134 laGluProProGlyLysProThrGlySerGluAspHisGlnHisGlyArg 1150
 3392 CAGAGCCCTCCGGAAGCCCAAGGATCAGAGACCATCAGCATGGCCGA 3441
 1151 AlaThrGlnLeuProGlyAlaLeuAspThrSerSerProGlyThrGlnH 1167
 3442 GCCACACAGCTCCAGAGCTGTGATACAAAGCTCCCAAGGAGCCGCA 3491
 1167 sProPheAlaProGluThrProIleProGlyAlaSerTrpSerIleSerP 1184
 3492 TCCCTTTCCTCCCTGAGACACCAATCCTGGAGCATCCTGGAGCATCTCC 3541
 1184 roThrThrProGlyLysLeuProTrpGlyTrpThrGlnThrProThrPro 1200
 3542 CTACACACCCCGGGGGGCTGCTGGGGCTGAGCTCAGACACCTACGCCA 3591
 1201 ValProGluAspLysGlyGlnProGlyLysLeuAspLeuArgHisProGly 1217
 3592 GTCCCTGAGGCAAAAGGCAACCTGGAGAGACCTGAGACATCCCGGCAC 3641
 1217 rSerLeuProAlaAspLeuPro 1224
 3642 CAGCCTCCCTGCTGCTGCCCG 3663

seq_name: gb-pat:AX319856

seq_documentation_block: 3561 bp DNA linear PAT 14-DEC-2001
 LOCUS AX319856
 DEFINITION Sequence 20 from Patent WO0183782.
 ACCESSION AX319856
 VERSION AX319856.1 GI:17901446

KEYWORDS

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (sites)
 Plowman, G.D., Whyte, D., Sudarsanam, S., Manning, G., Caenepeel, S. and

Payne, V.
 Novel proteases
 Patent: WO 0183782-A 20 08-NOV-2001.
 Sugen, Inc. (US)

FEATURES
 source
 1. 3561
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 694 a 1131 c 1113 g 623 t
 ORIGIN

alignment_scores:
 Quality: 6508.00 Length: 1224
 Ratio: 5.511 Gaps: 1
 Percent Similarity: 96.487 Percent Identity: 96.324

alignment_block:

US-09-938-330-20 x AX319856 ..

Align seg 1/1 to: AX319856 from: 1 to: 3561

1 MetAlaProLeuArgAlaLeuLeuSerTyrLeuLeuProLeuHisCysAl 17
 1 ArgGCTCCACCTCCGGCGCTGCTGTCTTACTGCTGCTTGCATCTGGC 50
 17 aleuCystrAlaAlaGlySerArgThrProGluLeuHisLeuSerGlyL 34
 51 GCTTCGCCCGCGCGGGGAGCGGAGCCCAAGAGCTGACCTCTCTGGA 100
 34 yalSerAspTyrGlyValThrValProCysSerThrAspPheArgGly 50
 101 ACCTCAGTCACTATGTGTGACAGTGCCTGACGACGACACTTCGGGGA 150
 51 ArgPheLeuSerHisValSerGlyProAlaAlaAlaSerAlaGlySe 67
 151 GCTTCTCTCTCCACGCTGCTGTCTGCTGCCAGAGAGAGCTCTGCGAG 200
 67 rMetValAlaSerThrProProThrLeuProArgHisSerSerHisLeu 84
 201 CATGGTAGTGACAGCCACCACACTACACAGACACTCCAGTCACTCC 250
 84 rGlyAlaAlaArgSerProLeuHisProGlyGlyThrLeuTrpProGlyArg 100
 251 GGGTGGCTGCGAGCCCTGACACCCAGAGGAGACCTGTGGCTGGCAG 300
 101 ValGlyArgHisSerLeuTyrPheAsnValThrValPheGlyLysGlu 117
 301 GTGGGGCGGCACCTCCCTCACTCAATGTCACCTGTTTCGGGAAGAACT 350
 117 whiLeuArgLeuArgProAsnArgArgLeuValValProGlySerSer 134
 351 GCACCTTGCCCTGGCGCCCAATCGGAGGTGGTAGTGCCAGGATCCAG 400
 134 alGluTrpGlnGluAspPheArgGluLeuPheArgGlnProLeuArgGln 150
 401 TGGAGTGGCAGAGAGATTTTCGGGAGCTGTCCGGAGCCCTTACGGCAG 450
 151 GluCyValTyrThrGlyGlyValThrGlyMetProGlyAlaAlaValAl 167
 451 GAGTGTGTGACACTGAGGTGCACTGGAATGCCGGGGCGAGCTGTTC 500
 167 alieSerAsnCysAspGlyLeuAlaGlyLeuLeuArgThrAspSerTh 184
 501 CATCAGCAACTGTGACGATGTGGGGCTCATCCGCAACAGACAGCAGCC 550
 184 sPhePheIleGluProLeuGluArgGlyGlnGlnLeuLysGluAlaSer 200
 551 ACTTCTTCATTTAGCCTCTGAGCGGGGGCCAGCAGAGAAGGCGCAGC 600
 201 GlyArgThrHisValValTyrArgArgGluAlaValGlnGlnGluTrpAl 217
 601 GGGAGGACACATGTGTGTACCCCGGGAGGCCGTCCACAGAGAGTGGGC 650
 217 agluProAspGlyAspLeuHisAsnGlnAlaPheGlyLeuGlyAspLeu 234
 651 AGAACCTGACGGGAGACTGCACAATGAAGCTTTGGCGCTGGAGACTTC 700
 234 roAsnLeuLeuGlyLeuValGlyAspGlnLeuGlyAspThrGluArgLys 250
 701 CCAACCTGTGGGCTGTGGGGAGCAGCTGGGGCAGACAGAGCGGAAG 750
 251 ArgArgHisAlaLysProGlySerTyrSerIleGluValLeuLeuVal 267
 751 CGGGCGCATGCCAAGCCAGGACCTCAGCATCAGAGTGTGCTGTGGT 800
 267 laspAspSerValValAlaArgPheHisGlyLysGluHisValGlnAsnTyr 284
 801 GGAGCACTCGGTGCTGCTTCATGCAAGAGAGCATGTGCAAACTATG 850
 284 alLeuThrLeuMetAsnIleValAspGluIleTyrHisAspGluSerLeu 300
 851 TCTCACCCCTCATGATATATGATGATGATTTTACACGATGATGCTCC 900

```

301 GlyValHisIleasnIlealValArgLeuIleMetValGlyTyr 317
    |||||||
901 GGGGTTCAATATAATATTCCTCTCCGCTTGATCATGTGGCTACCG 950
    |||||||
317 gGlnSerLeuSerLeuIleGlyArgGlyAsnProSerArgSerLeuG 334
    |||||||
951 ACAGTCCCTGAGCCTGATGAGACGGGGACCCCTCAGCAGCCTGGAGC 1000
    |||||||
334 lValCysArgTrpAlaHisSerGlnIleArgGlnAspProSerHisAla 350
    |||||||
1001 AGGTGTGTCGCTGGGACACATCCAGCAGCGCCAGACCCAGCAGCCT 1050
    |||||||
351 GlnHisHisAspHisValValPheLeuThrArgGlnAspPheGlyProse 367
    |||||||
1051 GAGCACCATGAGCAGCTGTGTCTCTACCCGGCAGAGACTTTGGGCTCTC 1100
    |||||||
367 rGlyTyrAlaProValIleThrGlyMetCysHisProLeuArgSerCysAla 384
    |||||||
1101 AGGGATGACCCCGTCACTGGCATGTGTACACCCCTGAGAGCTGTGCC 1150
    |||||||
384 euAsnHisGlnAspGlyPheSerSerAlaPheValIleAlaHisGlyThr 400
    |||||||
1151 TCACCATGAGAGATGGCTTCTCTCAGCCTTGATATAGCTCATGAGAGC 1200
    |||||||
401 GlyHisValLeuGlyMetGlnHisAspGlyGlnGlyAsnGlyCysAlaAs 417
    |||||||
1201 GGCCACGCTGCTGGCATGAGCATGAGCAGGTCAAGGAAATGGCTGTGAGA 1250
    |||||||
417 pGluThrSerLeuGlySerValIleMetAlaProLeuValGlnAlaAlaPhe 434
    |||||||
1251 TGAGCACGAGCTGGGACAGCTCATAGGCGCCCTGTGTCAGGCTGTCC 1300
    |||||||
434 lArgPheHisTrpSerArgCysSerLysLeuGlnLeuSerArgTyrLeu 450
    |||||||
1301 ACCGCTTCATTTGGTCCCGCTGACAGCAAGTGAGCTCAGCCCTACCTC 1350
    |||||||
451 ProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTrpProG 1400
    |||||||
1351 CCTCTCTAGGACTGCTCTGTGATGACCCCTTGTATCTCTCTGCGGCCCA 1450
    |||||||
467 nProProGluLeuPProGlyIleAsnTyrSerMetAspGlyGlnCysArgP 484
    |||||||
1401 GCCCCAGAGCTGCCGTGGATCAACCTCACTGATGAGATGGACATGGCGCT 1450
    |||||||
484 heAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheGln 500
    |||||||
1451 TTGACTTTGGCAGTGGCTACACAGACCTGTGGCATTCAGAGACTTTGAG 1500
    |||||||
501 ProCysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCysLys 517
    |||||||
1501 CCTGCAAGCAGAGCTGTGTGTCAGACCATCTGACACACCCGATCTTGCA 1550
    |||||||
517 sThrLysLysGlyProProLeuAspGlyThrGlyCysAlaProGlyLysT 534
    |||||||
1551 GACCAAGAGAGGGGCCCCGCTGGATGGAGTGTGATGTCACCCGGCAAT 1600
    |||||||
534 rPysPheLysGlyHisCysIleTrpLysSerProGlnIleThrTyrGly 550
    |||||||
1601 GGTCCTCAAGGTCACATGATCTGGAAGTCGCGGAGCAGACATATGAC 1650
    |||||||
551 GlnAspGlyLysTrpSerSerTrpThrLysPheGlySerCysSerArgse 567
    |||||||
1651 CAGGATGAGAGCTGGAGCTCTGACCAAGATTGGGTCTGTTCCGGGGCTC 1700
    |||||||
567 rGysGlyGlyGlyValAlaArgSerArgSerArgSerCysAsnAsnProSerP 584
    |||||||
1701 ATGTGGGGGGGGGGTGGATCCCGAGCCGAGACTGCACAAACCCCTGCC 1750
    |||||||
584 roAlaTyrGlyGlyArgProCysLeuGlyProMetPheGlnTyrGlnVal 600
    |||||||
1751 CAGGCTATGAGGCGCCGCTGTAGGGCCCATGTTCAGATACAGAGGC 1800
    |||||||
601 CysAsnSerGlnGluCysProGlyThrTyrGlnAspPheArgAlaGlnG 617
    |||||||
    |||||||
1801 TCACAGGAGGAGAGTGGCTGGGACTACAGAGACTCCGGGCCAGCA 1850
    |||||||
617 nCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSerT 634
    |||||||
1851 GTGTGCCAAGCCCAATGCTACTATGTGTACCAAGAAATGCCAAGCAGCT 1900
    |||||||
634 rPValProTyrGluProAspAspAlaGlnLysCysGluLeuIleCys 650
    |||||||
1901 GGGTGCCCTACAGAGCCTGACGATGACGCCCAAGAAAGTGTGACTATCTGC 1950
    |||||||
651 GlnSerAlaAspThrGlyAspValValPheMetAsnGlnValValHisAs 667
    |||||||
1951 CAGTGGGGGACACGGGGAGCTGTGTTCATGAACCAAGAGGTTCACGA 2000
    |||||||
667 pGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArgGly 684
    |||||||
2001 TGGACACGCTGACAGCTACCGGAGCCATACAGCGCTGTGCGCGTGGCG 2050
    |||||||
684 lCysValProValGlyCysAspLysGlnValGlySerMetLysAlaAsp 700
    |||||||
2051 AGTGTGCTGCTGGCTGGCTGTGACAAAGAGGTGGGGTCCATGAAGCGGAT 2100
    |||||||
701 AspLysCysGlyValCysGlyLysAspAsnSerHisCysArgThrValLys 717
    |||||||
2101 GACACAGTGGAGTGTGGGGGGGTGACACCTCCACCTGACAGACTGTGA 2150
    |||||||
717 sGlyThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuValG 734
    |||||||
2151 GGGGAGCTGGGCAAGGGCTCCAAAGCAGGACGAGCTCAACCTGTGTG 2200
    |||||||
734 lIleProAlaGlyAlaArgHisIleGlnIleGlyAlaLeuGlnLysSer 750
    |||||||
2201 AGATCCACAGCAGGTGCCAGGACATCCAGATTGAGGACACGAGGAAGTCC 2250
    |||||||
751 ProHisArgSerValValLysAsnGlnValThrGlySerPheIleLeuAs 767
    |||||||
2251 CCCCACGCAATTGTGTGGAAGAACACAGCTACCGGACGCTTACTCTCAA 2300
    |||||||
767 nProLysGlyLysGlnAlaThrSerArgThrPheThrAlaMetGlyLeuG 784
    |||||||
2301 CCCCAGGGGCAAGGAGAACCCAGACCGGACCTTACCGCATGGGCTGG 2350
    |||||||
784 lIuTrpGluAspAlaValGluAspAlaLysGlnSerLeuLysTrpSerGly 800
    |||||||
2351 AGTGGAGAGATGCCGCTGAGAGATGCCAAGAAAGCCTCAAGACACCGGG 2400
    |||||||
801 ProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGlnGly 817
    |||||||
2401 CCTCTGCTGAAGCATTTGCCATCTGTGCTTCCCCCAACTGAGGGTGG 2450
    |||||||
817 yProArgSerSerLeuAlaTyrTrpTyrValIleHisGluAspLeuLeuP 834
    |||||||
2451 CCCCAGGAGAGCCTGGCTGACAAAGTACGTACCTACCTGAGGAGCTGTGC 2500
    |||||||
834 roLeuIleGlySerAsnAsnValLeuLeuGlnGlnMetAspThrTyrGlu 850
    |||||||
2501 CCTTATCGGAGACAACAATGTCTCTCTGAGAGAGATGACACCTATGAG 2550
    |||||||
851 TrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGlyLys 867
    |||||||
2551 TGGCGCTCAAGAGCTGGGGCCCCCTGACCAAGGCTGTGAGGAGGAGAT 2600
    |||||||
867 eGlnPheThrLysTyrGlyCysArgArgArgArgAspHisHisMetValG 884
    |||||||
2601 CCAAGTTCACCAATACGGCTGCGGGGCGAGAGAGACCAACCATGAGTGC 2650
    |||||||
884 lAlaArgHisLeuCysAspHisLysArgProLysProIleArgArgArg 900
    |||||||
2651 AGCGACACCTGTGTGACCAAGAAAGGCCCAAGCCCATCGCGCGGGC 2700
    |||||||
901 CysAsnGlnHisProCysSerGlnProValIleThrGluGluTrpGly 917
    |||||||

```


351 GCACTTGCCCTGGGGCCCAATCGAGGTTGTAGTCCAGCATCTCTCAG 400
 134 ALGIUTPGLINGluAspPheArgGluLeuPheArgGluProLeuArgGln 150
 401 TGGAGTGGCAGAGAGATTTTCGGAGCTGTCTCGGACGCCCTTACGGCAG 450
 151 GlucysValTyrThrGlyGlyValThrGlyMetProGlyAlaAlaValAla 167
 451 GAGGTGTGTACACTGGAGGTGTACATGGAAATGCTGGGGCAGCTGTTC 500
 167 AlIeSerAsnCysAspGlyLeuAlaGlyLeuIleArgThrAspSerThr 184
 501 CATCAGCAACTGTGACGAGATGGCGGGCTCATCCGACACAGACAGACCG 550
 184 sPhePheIleGluProLeuGluArgGlyGlnGlnGluLysGluLaser 200
 551 ACTTCTTCATTGAGCTCTGGAGCGGGCGCCAGCAGAGAGAGAGGCGCAGC 600
 201 GlyArgThrHisValValTyrArgArgGluAlaValGlnGlnGluTrrAl 217
 601 GGGAGGACACATGTGGTGTACCGCGGGAGCGCTCCAGCAGAGATGGGC 650
 217 agIuProAspGlyAspLeuHisAsnGluAlaPheGlyLeuGlyAspLeuP 234
 651 AGAACCTGACGGGAGCTGCACAAATGAACCTTGGCTGGGAGACCTTC 700
 234 roAsnLeuLeuGlyLeuValGlyAspGlnLeuGlyAspThrGluArgLys 250
 701 CCMACTGCTGGGCCCTGGTGGGGGACCACTGGCGACACAGAGGGGAG 750
 251 ArgArgHisAlaLysProGlySerTyrSerIleGluValLeuLeuValVa 267
 751 CGCGCGCAATGCCAGCAGCAGCTACAGCATCGAGTGTGCTGTGTGT 800
 267 LaspAspSerValValArgPheHisGlyLysGlnHisValGlnAsnTyrV 284
 801 GAGCAGCTGGGTGGTTCCTTCATGTCAGAGAGCATGTGCAGAACTATG 850
 284 alIeuthrLeuMetAsnIle...ValAspGluIleTyrHisAspGluSer 299
 851 TCCTCACTCCATGAATATCGTGTAGATGAGATTACACAGATGAGATCC 900
 300 LeuGlyValHisIleAsnIleAlaLeuValArgLeuIleMetValGlyTyr 316
 901 CTGGGGGTTCAATATAATATGCTCGCTCGCTGATCATGATGTGGCTA 950
 316 rArg...GlnSerLeuSerLeuIleGluArgGlyAsnProSerArgSerL 332
 951 CGGACAGCAGTCCCTGAGCTGATCGAGCGGGGAACCCCTCACGCGACCC 1000
 332 euGluGlnValCysArgTrrAlaHisSerGlnArgGlnAspProSer 348
 1001 TGGACAGAGTGTGTGGCGGACACTCCAGCGCGCCAGAGACCCAGC 1050
 349 HisAlaGlnHisHisAspHisValValPheLeuThrArgGlnAspPheG 365
 1051 CAGGCTGAGCACCATGACACAGCTGTGTCTCTACCCGCGAGGACTTGG 1100
 365 yProSer...GlyTyrAlaProValThrGlyMetCysHisProLeuArg 381
 1101 GCCCTCAGGTGGTATGACACCCGTACATGCTGTGCACCCCTTGAGGA 1150
 381 eCyAlaLeuAsnHisGluAspGlyPheSerSerAlaPheValIleAla 397
 1151 GCTGTGCCCTCAACCATGAGGATGGCTTCTCCACGCTTCGATAGT 1200
 398 HisGluThrGlyHisValLeuGlyMetGlnHisAspGlyGlnGlnAsnG 414
 1201 CATGAGACCGGCGCACGTGCTGTGGCATGAGCATGACGTCAGGGGATGG 1250
 414 yCysAlaAspGluThrSerLeuGlySerValMetAlaProLeuValGlnA 431
 1251 CTGTGACAGATGAGACAGCTGGGCGACGTCATGGCGCCCTGTGTGACAG 1300
 431 lAlaPheHisArgPheHisTrrSerArgCysSerLysLeuGluLeuSer 447
 1301 CTGGCTTCCACCGCTTCATGTGTTCCCGCTGCACAGCATGAGCTCAGC 1350
 448 ArgTyrLeuProSerTyrAspCysLeuLeuAspAspProPheAspProAl 464
 1351 CGCTACCTCCCGTCTACAGATGCTCTCGATGACCCCTTGTATGCTGCG 1400
 464 aTrpProGlnProGluLeuPheGlyIleAsnTyrSerMetAspGluG 481
 1401 CTGGCCCGACCCCGCAGACTGCTGTGGATCAACTCAATCAATGATGAGC 1450
 481 lncysArgPheAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArg 497
 1451 AGTGGCGCTTGTACTTGTGGCAGTGGCTACACAGACCTGTGTGCATTCAG 1500
 498 ThrPheGluProCysLysGlnLeuTrrPcySerHisProAspAsnProTy 514
 1501 ACCTTGTAGCCCTGCAGAGAGCTGTGTGACAGCCATCTGTACAAACCCGTA 1550
 514 rPheCysLysThrLysLysGlyProProLeuAspGlyThrGluCysAlaP 531
 1551 CTTCCTCAAGACCAAGAAAGGGCCCCCGCTGGATGGAGCTAGTGTGCAC 1600
 531 roGlyLysTrrPcyPheLysGlyHisCysIleTrrPlySerProGluGln 547
 1601 CCGGCAAGTGTGCTTCAAAAGCTCACTGCATCTGCAATGCCCGAGACAG 1650
 548 ThrTyrGlyGlnAspGlyGlyTrrPserSerTrrPrrLysPheGlySerCy 564
 1651 ACATATAGGCGCAGATGAGAGCTGTGAGCTCTGGACCAAGTTGGAGTATG 1700
 564 sSerArgSerCysGlyGlyValValArgSerArgSerArgSerCysAsnA 581
 1701 TTGGGGGTATGTGGGGGGGGGTGCATCCCGACGCGGAGCTGCACAA 1750
 581 snProSerProAlaTrrGlyGlyArgProCysLeuGlyPrometPheGlu 597
 1751 ACCCTCGCGCAGCTATGAGGCGCCCTGTGTAGGGCCCATGTTCGAG 1800
 598 TyrGlnValCysAsnSerGluGluCysProGlyThrTyrGluAspPheAr 614
 1801 TACCAAGTGTGCACAGGAGAGATGCCCTGGACCTACAGAGACTTCGG 1850
 614 gAlaGlnGlnCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaL 631
 1851 GGCCGAGAGTGTGCCAAGCGCAACTCTACTATGTGCACACAGAAATGCCA 1900
 631 yHisSerTrrValProTrrGluProAspAspAlaGlnLysCysGlu 647
 1901 AGCAGAGCTGGGTGCCCTTACAGGCTTACAGATGACGCCCAAGAGTGTGAG 1950
 648 LeuIleCysGlnSerAlaAspThrGlyAspValValPheMetAsnGlnVa 664
 1951 CTGATCTGCCAGTGGGGGAGACGGGGGAGCTGTGTCTTGAACCAAGGT 2000
 664 lValHisAspGlyThrArgCysSerTyrArgAspProTrrSerValCysA 681
 2001 GGTTCACGATGGGACAGCGTGCACGCTACCGGAGCCATACAGCGTGTG 2050
 681 lAArgGlyGluCysValProValGlyCysAspLysGluValGlySerMet 697
 2051 CGCGTGGCGAGTGTGTGCTGTGGCTGTACAAAGAGGTGGGGTCCAG 2100
 698 LysAlaAspAspLysCysGlyValCysGlyGlyAspAsnSerHisCysAr 714
 2101 AAGCGGATGACAAAGTGTGAGTCTGGGGGGTGAACAACATCCACATCGAG 2150
 714 gThrValLysGlyThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeu 731
 2151 GACTGTGAAGGGAGCGCTGGGCAAGGCTTCCAAAGCAGGAGGTGTCTCA 2200

731 ysleuValGlnIleProAlaGlyAlaArgHisIleGlnIleGlnAlaLeu 747
 2201 AGCTGTGTCAGATCCACAGCAGGTGCGACATCCAGATTGAGGCACTG 2250
 748 GlnUysSerProHisArgSerValValLysnGlnValThrGlySerPh 764
 2251 GAGAGTCCCTCCACCGATTGTGTGAAGAACAGGATCCAGCCGAGCTT 2300
 764 eileuAnsProLysGlyLysGlnAlaThrSerArgThrPheThrAlaM 781
 2301 CATCTTCACACCCCAAGGCGCAAGAACCAAGCCGAGCTTCACCGCA 2350
 781 etGlyLeuGlnTrpGlnAspAlaValGluAspAlaLysGlnSerLeuLys 797
 2351 TGGGCTGAGTGGAGGATGCGGTGGAGATGCCAAGAAAGGCTCTCAG 2400
 798 ThrSerGlyProLeuProGlnAlaIleAlaIleAlaLeuAlaLeuProProTh 814
 2401 ACCAGCGGGCTCTGCTGAGAACCTTCATCTGCTCTCCCTCCAC 2450
 814 rGlnGlyGlyProArgSerSerLeuAlaIleLysTrpValIleHisGln 831
 2451 TGAGGTGGCCCCCAGCAGAGCTGCTACAGATCCTCATTCATGAG 2500
 831 spleuLeuProLeuIleGlySerAsnAsnValLeuLeuGlnGluMetAsp 847
 2501 ACCCTGCTGCTTATCGGAGCAACATGTCTCTTGAGAGATGAGAC 2550
 848 ThrTrpGlnTrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGln 864
 2551 ACCTATGATGGGCGCTCAAGAGCTGGGCCCCCTCAGCAAGGCTGTGG 2600
 864 yGlyGlyIleGlnPheThrLysTrpGlyCysArgArgArgArgAspHis 881
 2601 AGGAGGTATCCAGTTCACCAATACGGCTGCGGCGCAGACAGACACC 2650
 881 iMetValGlnArgHisLeuCysAspHisLysLysArgProLysProIle 897
 2651 ACATGTGTGAGGAGACACCTGTGTGACCAACAGAGGCGCCACAGCCAT 2700
 898 ArgArgArgCysAsnGlnHisProCysSerGlnProValTrpValThrl 914
 2701 CGCGGAGCTGCAACAGCACCGCTCTCTCAGCTGTGTGGGTACGGA 2750
 914 uGlnTrpGlyAlaCysSerArgSerCysGlyLysLeuGlyValGlnThra 931
 2751 GAGGTGGGTGCTGCAAGCGGAGCTGTGGAAAGCTGGGGGTGCAAGAC 2800
 931 rGlyIleGlnCysLeuLeuProLeuSerAsnGlyThrHisLysValMet 947
 2801 GGGGATATACAGTGTGCTGCTCTCTCCAAATGAAACCCCAAGAGTCAT 2850
 948 ProAlaLysAlaCysAlaGlyAspArgProGlnAlaArgArgProCysLe 964
 2851 CCGGCCAAAGCTGCGCGGAGGAGCGGCTGAGGCGCCGAGCGGCTGT 2900
 964 uArgValProCysProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCys 981
 2901 CGAGTGGCTGCTGCGCAGCCAGTGGAGCTGGAGGCTGTGCTCCAGTGT 2950
 981 eAlaThrCysGlyGlyIleGlnGlnArgGlnValValCysArgThr 997
 2951 CTGCCACCTGTGTGAGGAGGATCCACAGCGGAGGTGTGTGTCCAGACC 3000
 998 AsnAlaAsnSerLeuGlyHisCysGlyGlyAspArgProAspThrValGln 1014
 3001 AAGGCCAAGAGCTCGGCAATTGCGAGGGGATAGCCAGACACTTCCA 3050
 1014 nValCysSerLeuProAlaCysGlyLysnHisGlnAsnSerThrValAla 1031
 3051 GGTCTGAGGCTGCGCGCTGTGA..... 3075
 1031 rGlnAspValTrpGlnLeuGlyThrProGlnGlyGlnTrpValProGln 1047

3075 3075
 1048 SerGlyProLeuHisProIleAsnLysIleSerSerMetCysAlaAlaGln 1064
 3076 GCGGA 3080
 1064 uProCysThrGlyAspArgSerValPheCysGlnMetGluValLeuAspA 1081
 3081 GCCCTGACGGAGACAGCTGTCTTCTTCCAGATGATGAAAGCTCGAGAC 3130
 1081 rGlyCysSerIleProGlyTrpHisArgLeuCysCysValSerCysIle 1097
 3131 GCTACTGCTCATTCCTCGGCTACACCGGCTGCTGTGTGTCTGCTGACAC 3180
 1098 LysLysAlaSerGlyProAsnProGlyProAspProGlyProThrSerle 1114
 3181 AGAAGGCTCGGGCCCAACCTGGCCAGACCTGGGCCCAACCTCACT 3230
 1114 uProProPheSerThrProGlySerProLeuProGlyProGlnAspProA 1131
 3231 GCCCTCTTCTCCACTCTCTGAGAGCCCTTACAGAGACCCAGGACCTG 3280
 1131 LaaspAlaAlaGluProProGlyLysProThrGlySerGlnAspHisGln 1147
 3281 CAGATGCTGCGAGCCTCTCGAAGCAAGGAGATCAGAGGACCATCAG 3330
 1148 HisGlyArgAlaThrGlnLeuProGlyAlaLeuAspThrSerSerProGln 1164
 3331 CATGGCCGAGCACACAGCTCCAGAGAGCTGTGATACAACTCCCAAG 3380
 1164 yThrGlnHisProPheAlaProGlnTrpProIleProGlyAlaSerTrps 1181
 3381 GACCACATCCCTTTGCCCCGTGAGACACCAATCCCTGGAGCATCTGGA 3430
 1181 eIleSerProTrpThrProGlyGlyLeuProTrpGlyTrpThrGlnTrp 1197
 3431 GCATCTCCCTCACACCCCGGGGGGTGCTGTGGGCTGAGATCTGAGACA 3480
 1198 ProThrProValProGluAspLysGlyGlnProGlyLysLeuAspThr 1214
 3481 CCTACGCGACATCCCTCGAGGAGCAAAAGGCAACCTGAGAGACCTGAGACA 3530
 1214 sProGlyThrSerLeuProAlaAspLeuPro 1224
 3531 TCCCGGACAGCCTCCCTGCTGCTCCCGG 3561
 Seq. name: gb_pat:AX211284

seq. documentation block:
 LOCUS AX211284 3144 bp DNA linear PAT 06-SEP-2001
 DEFINITION Sequence 1 from Patent WO0159133.
 ACCESSION AX211284
 VERSION AX211284.1 GI:15523693
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 3144)
 AUTHORS Duecker,K., Heitsch,B., Hoheisel,J. and Frohme,M.
 TITLE Larynx carcinoma-associated protein larcap-1
 JOURNAL Patent: WO 0159133-A 1 (16-AUG-2001)
 MERCK PATENT GmbH (DE)
 FEATURES
 source location/Qualifiers
 1..3144
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..>3144
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC69262.1"
 /db_xref="GI:15523694"


```

/translaction="MHAASAHSPVLFLMLPWLAKGSOVILLILOGPCVCLIEECPL
LPQGRGVGGTLLGOMOLERYTGMLGAGGSGISQDEYTHHDESLCVHINIALVRLIM
VGRYSILIERGNSRSLRBOYGRMAHSQOROPSHAEHHDHVFLTRDFGSGTAP
VTGMCPLRSLALNHEDESSAFYIAHETGVLHMGEDGNGCADETSIGSMALV
QAARHTHMSRCSKLELSRYLPSIDCLDDPFPAMPQPELPGIAYSMEDOCREFPG
SGYOTCLAFRTPEPCQJLWCSPDNPFCKTKKGPPLDGECAAPKWCRCIMKPS
EOTYQDQGWMSWTKFGSGSRSCGGVRSRSCGNPSPDGPGLDGLGMPFYOVNS
EECPGYEDFRAOQCAKRSYVYHONAKHSMVNEFPDDAKOCELIQOASADIDVYPM
NOVYDGTCSYDRPYSVCARCEVPCDCKEYSGKADKCGVCGDNSHCRTYKGT
LGKASKAGALKIVQIPAGARHIOLEALEKSPHRIYKNOVTSFTLNPKEATSRIT
FTAGLEWEDAVEDEKESLTSGLPEPALIALIPTEGGPRRSLAYKXVIEDLLPL
IGSNVNLLEMDYETEMALIKSMAPCSKACGGGIGQFTYCGRRRHHMVRHICDKKR
PKPIRRRCNOHPCSOYVWTVEWAGSRSRSCGGVRSRSCGNPSPDGPGLDGLGMPFYOVNS
CAGDRPEARPPCLRPVPCPAOWRLGAMSOKYILSTSCMDPLVLRMPRPSINOTELLAL
VOPTVLCATCGEIOOROVYCRTNANSIHCHEGDRPDTVOYCSLPCGNGHONSTVR
ADWELCTPEGQWYPOSEPLPHIINKISSTPCGDSVCOEVLDRYCSITGYHRLC
CVSCITKASSPNPBPDPGPTPLPSTPGSPDPDPAADAEPRGKPGSDHOGHR
ATDLPGLDTSSTPGTOHPAPETPIPGASMSISPTPGGLPMGWTQTPPVPEDEKGP
GEDLRHPTSLPASPVT"
BASE COUNT      633 a      1000 c      957 g      554 t
ORIGIN

```

```

alignment_scores:
    Quality: 5160.50      Length: 970
    Ratio: 5.573          Gaps: 2
    Percent Similarity: 95.464      Percent Identity: 95.361

```

alignment block:

US-09-938-330-20 x AX211284 ..

Align seg 1/1 to: AX211284 from: 1 to: 3144

```

291 ValAspGluIleTyrHisAspGluSerLeuGluValHisIleAsnIleAl 307
|||||
238 GTAAGATGAGATTACACGATGAGTCCCTGGGGGTCAATATAATATGTC 287
|||||
307 aLeuValArgLeuIleMetValGlyTyrArgGlnSerLeuSerLeuIleG 324
|||||
288 CCTCTCCGCTGATCATGTTGGGTACGACGACTCCCTGAGCCTGATCG 337
|||||
324 IuArgGlyAsnProSerArgSerLeuGluGlnValCysArgTTPAlaHis 340
|||||
338 AGCGGGGAGACCCCTCACGACGCTGAGGAGCGGTCTCCCTGGGCGCAC 387
|||||
341 SerGlnGlnArgGlnAspProSerHisAlaGlnHisAspHisValVa 357
|||||
388 TCCCAGACGCGCCAGAGACCCGACGCTGAGCATGACCATGACACGTTGT 437
|||||
357 IPhleuThrArgGlnAspPheGlyProSerGlyTyrAlaProValThrG 374
|||||
438 GTTCTCTACCCGCGAGGACTTGGGCTCAGGGTATGACACCGGTCACTG 487
|||||
374 IYMeCysHisProLeuArgSerCysAlaLeuAsnHisGluAspGlyPhe 390
|||||
488 GCAATGTGTACACCCCTGAGAGAGCTGTGCTCAACCATGAGAGATGGCTTC 537
|||||
391 SerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGluMetG 407
|||||
538 TCCTCAGGCTTCGTGATAGCTCATGAGACGCGCACCTGCTGGCATGGA 587
|||||
407 uHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuGlySerY 424
|||||
588 GCATGAGCGGTCAGGGAATGGCTGTGAGATGAGACGACCTGGGACAGG 637
|||||
424 aMetAlaProLeuValGlnAlaAlaPheHisArgPheHisIleSerArg 440
|||||
638 TCATGGGGCCCTGTGTGTCAGGCTGCTCCACCGCTTCATTTGGTCCGCG 687
|||||
441 CysSerIysLeuGluLeuSerArgTyrLeuProSerTyrAspCysLeu 457
|||||
688 TGCAGCAAGCTGAGAGCTCAGCGCTACCTCCCTCTACGAGCTGCTCTCT 737
|||||

```

```

457 uAspAspProPheAspProAlaTyrProGlnProProGluLeuProGlyI 474
|||||
738 CGATGACCCCTTTGATCTCGCTGGGCCACAGCCCAAGAGCTGCTGGGA 787
|||||
474 IeAsnTyrSerMetIAspGluGlnCysArgPheAspPheGlySerGlyTyr 490
|||||
788 TCACACTACTCAATGATGAGAGAGTCCGCTTGTGACTTTGGCAGTGCCTAC 837
|||||
491 GlnThrCysLeuAlaPheArgThrPheGluProCysIysGlnLeuTyrPcy 507
|||||
838 CAGACCTGCTTGGCATTTGAGACCTTTGAGCCCTTCAGACGACTGTGTTG 887
|||||
507 sSerHisProAspAsnProTyrPheCysIysThrIysIysGlyProProL 524
|||||
888 CAGCCATCTCTGACAAACCCGTACTCTGCAAGACCAAGAAAGGGGCCCGCC 937
|||||
524 euAspGlyThrGlyCysAlaProGlyIysTyrPcyPheGlyHisCys 540
|||||
938 TGGATGGACTGATGTTGACCCGGCAAGTGTGCTTCAAGGTCACCTGC 987
|||||
541 IleTyrPlySerProGluGlnThrTyrGlyGlnAspGlyGlyTyrPser 557
|||||
988 ATCTGGAAGCTCGCCGAGACGATATGCGAGATGAGAGGCTGAGACTC 1037
|||||
557 rTTPhrIysPheGlySerCysSerArgSerCysGlyGlyValArgS 574
|||||
1038 CTGGACCAAGTTGGGTCATGTCGGGTCAATGTGGGGGCGGGGTCCGAT 1087
|||||
574 eArgSerArgSerCysAsnAsnProSerProAlaTyrGlyGlyArgPro 590
|||||
1088 CCGCGACGCGAGAGCTGCAACAAACCCCTCCACGCTTATGGAGGCGGCTG 1137
|||||
591 CysLeuGlyProMetLeuPheGluTyrGlnValCysAsnSerGluCysPr 607
|||||
1138 TECTTAGGGCCCATGTTGAGTACAGAGTGTGCAACAGCGAGAGTGTCCC 1187
|||||
607 oGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaIysArgAsnSer 624
|||||
1188 TGGGACCTACAGAGACTTCGCGGCCCAAGCACTGTGCAACCGCAACTCCT 1237
|||||
624 YrTyrValHisGlnAsnAlaIysHisSerTyrValProTyrGluProAsp 640
|||||
1238 ACTATGTGCACAGATATGCCAAGCAGACGTGGGTGCTTACAGACCTGAC 1287
|||||
641 AspAspAlaGlnIysCysGluLeuIleCysGlnSerAlaAspThrGlyAs 657
|||||
1288 GATGACGCCCAAGAGTGTGAGCTGATCTGCCAGTGGCGGACACGGGGGA 1337
|||||
657 PValValPheMetAsnGlnValValHisAspGlyThrArgCysSerTyr 674
|||||
1338 CGTGGTGTTCATGAACAGGTGGTTCACGATGGGACACGCTGACGTACC 1387
|||||
674 rGAspProTyrSerValCysAlaArgIysGluCysValProValGlyCys 690
|||||
1388 GCGACCCATACAGGCTGTGCGGGTGGCGAGTGTGCTGCTGGCTGT 1437
|||||
691 AspIysGluValGlySerMetIysAlaAspAspIysCysGlyValCysG 707
|||||
1438 GACAAGAGAGGTGGGTCCATGAAGCGGATGACAAAGTGTGAGATGCGCG 1487
|||||
707 YGlyAspAsnSerHisCysArgThrValIysGlyThrLeuGlyIysAla 724
|||||
1488 GGGTGACAACTCCACCTGCGAGACTGTGAAGGGGAGAGCTGGGGAGGGCT 1537
|||||
724 eTlySGlnAlaGlyAlaLeuIysLeuValGlnIleProAlaGlyAlaArg 740
|||||
1538 CCAAGCAGCAGAGAGCTCTCAAGCTGTGACAGATCCACAGAGTGCAGG 1587
|||||
741 HisIleGlnIleGlnAlaLeuGluIysSerProHisArgSerValVally 757
|||||
1588 CACATCCAGATTAGGACTGGAGAAAGTCCCCCACCAGATTGTGGTGA 1637
|||||
757 saGlnValThrGlySerPheIleLeuAsnProIysGlyIysGluAlaT 774
|||||

```

1638 GAACAGAGTACCGGAGCTTCATCCCAACCCCAAGGSCAAGAACCA 1687
 774 hrSerArgThrPheThrAlaMetGlyLeuGluTrpGluAspAlaValGlu 790
 1688 CAAAGCCGAGCTTCACCGCCCTGGGCTGAGTGGAGGATGGCGGTGAG 1737
 791 AspAlaLysGluSerLeuLysThrSerGlyProLeuProGluAlaIleal 807
 1738 GATGCCAAGAAAGCCTCAAGACAGCGGGCCCTGCTCAAGCCATATGC 1787
 807 alleLeuAlaLeuProProthTrpGluGlyGlyProArgSerSerLeuAla 824
 1788 CATCTGCTGCTCTCCCGCACTGAGGGGGCCCCCGCAGCAGCCTGGCCT 1837
 824 yTrpTyTrValIleHisGluAspLeuLeuProLeuIleGlySerAsnAsn 840
 1838 ACAAGTACGTATCATGAGAGCCTGCTGCCCTTATCGGAGCAACAAT 1887
 841 ValLeuLeuGluGluMetAspThrTrpGluTrpAlaLeuLysSerTrpAl 857
 1888 GTGCTCTGGAGAGATGAGACACTATGATGGGGCTCAAGAGCTGGGC 1937
 857 aProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTyrglyc 874
 1938 CCCCTGCAAGCAAGCCTGGAGGAGGATCAGTTCAACAATAACGCT 1987
 874 ySArGArGArGArAspHisSHMetValGlnAtrHisLeuCysAspHis 890
 1988 GCCGGCGCAGAGAGACCAACCATGATGAGCGACACCTCTGTGACAC 2037
 891 LysLysArgProLysProIleArgArgArgCysAsnGlnHisProCysSe 907
 2038 AAGAAAGAGCCCAAGCCCATCCGGCGGCTGCACCAAGCAACCGTGGTC 2087
 907 rGlnProValTrpValThrGluGluTrpGlyAlaCysSerArgSerCysG 924
 2088 TCAGCCTGTGGTGAGAGAGAGTGGGGTGCCTGCAGCCGAGCCTGTG 2137
 924 yLysLeuGlyValGlnThrArgGlyIleGlnCysLeuLeuProLeuSer 940
 2138 GGAAGCTGGGGGTGCAGACAGGGGATACAGTGTCTGCTGCTCTCC 2187
 941 AsnGlyThrHisLysValMetProAlaLysAlaCysAlaGlyAspArgPr 957
 2188 AATGGAAACCAAGGTATGCGCGGCAAAAGCCTGCGCGGAGCGGCGC 2237
 957 oGluAlaArgArgProCysLeuArgValProCysProAlaGlnTrpArgL 974
 2238 TGAAGGCCAGCGCCCTGTCTCGAGTGCCTGCGCAGCCAGTGAAGGC 2287
 974 euGlyAlaTrpSerGln..... 979
 2288 TGGAGAGCTGTGTCCAGAAATATTCTGTAGACATAGCTGTATGCCAGAC 2337
 979 979
 2338 CTGTACTAAGATGAGGAACCTAGCATCAACAGACAGAGCTGATTCT 2387
 980 CysSerAlaThrCysGlyGlyIle 988
 2388 TGCCCTCGTGACGCCACAGTCTTGTGCTGTGACACTGTGGAGAGGGA 2437
 988 leGlnGlnArgGlnValLysArgTrpAsnAlaAsnSerLeuGlyHis 1004
 2438 TCCAGCAGCGGCGAGGTGTGTGAGACCAACAGCCCAAGCCTCGGAGAT 2487
 1005 CysGluGlyAspArgProAspThrValGlnValCysSerLeuProAlaCys 1021
 2488 TGGAGAGGGGATAGGCCAGACACTGTCCAGGTCTGCAAGCCTGCGCCCTG 2537
 1021 sGlyGlyAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLeu 1038

2538 TGGAGAAATCACCAAGATCCACGCTGAGGCGCATGTCTGGAACTTG 2587
 1038 yThrProGluLysGlnTrpValProGlnSerGlyProLeuHisProIle 1054
 2588 GAGCGCCAGAGGGGAGGTGGGTGCCACATCTGAACCCCTTACATCCATT 2637
 1055 AsnLysIleSerSerMetCysAlaIleGluProCysTrpGlyAspArgSe 1071
 2638 AACAAATATCATCA.....ACGAGCCCTGACAGGAGACAGGCTC 2678
 1071 rValPheCysGlnMetGluValLeuAspArgTrpCysIleProGlyTrp 1088
 2679 TGTCTTCTCCAGATGGAAGTCTGATGCTGATGCTGATGCTGATGCTG 2728
 1088 yTrpHisArgLeuCysCysValSerCysIleLysLysAlaSerGlyProAsn 1104
 2729 AACACCGGCTCTGCTGTGTCTGCTGATCAAGAGGCTCGGGGCCCAAC 2778
 1105 ProGlyProAspProGlyProThrSerLeuProProPheSerThrProG 1121
 2779 CTGGCCCAAGACCCCTGGCCCAACCTCAGTCCCGCCCTTCTCAGTCTGG 2828
 1121 ySerProLeuProGlyProGlnAspProAlaAspAlaIleGluProProG 1138
 2829 AAGCCCTTACAGAGACCCAGAGACCTGACATGCTGAGAGGCTCTGG 2878
 1138 LysProThrGlySerGluAspHisGlnHisGlyValArgAlaThrGlnLeu 1154
 2879 GAAAGCCCAAGGAGATCAGAGACCATCAGCATGGCGGACCAACAGCTC 2928
 1155 ProGlyAlaLeuAspThrSerSerProGlyThrGlnHisProPheAlaPr 1171
 2929 CAGAGAGCTTGATACAGCTCCCAAGGAGACCCAGACATCCCTTTCGCC 2978
 1171 oGluThrProIleProGlyAlaSerTrpSerIleSerProThrTrpProG 1188
 2979 TGAAGACCAATCCCTGAGCATCTGGAGCATCTGCCATCCACACCCCG 3028
 1188 yGlyLeuProTrpTrpGlyTrpThrGlnThrProThrProValProGluAsp 1204
 3029 GGGGGCTGCTTGGGGCTGAGCTGAGACCTAGACCTAGCCAGTCCCTGAGAG 3078
 1205 LysGlyLysProGlyGluAspLeuArgHisProGlyThrSerLeuProAl 1221
 3079 AAGGGCAACCTGAGAGAGCTGAGACATCCCGGACAGCCTCCCTGCG 3128
 1221 aAspLeuPro 1224
 3129 TGCCCTCCCG 3138
 seq_name: gb_pat:E58661
 seq_documentation_block:
 LOCUS E58661 3618 bp DNA linear PAT 31-JAN-2002
 DEFINITION Novel metallic protease.
 ACCESSION E58661
 VERSION E58661.1 GI:18629883
 KEYWORDS JP 2001017183-A/9.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3618)
 AUTHORS YamaJ,I,N., Nishimura,K. and Sasamata,M.
 TITLES Novel metallic protease
 JOURNAL Patent: JP 2001017183-A 9-23-JAN-2001;
 YAMANOUCHI PHARMACEUT CO LTD
 COMMENT OS Homo sapiens (human)
 PN JP 2001017183-A/9
 PD 23-JAN-2001
 PF 09-JUL-1999 JP 1999196584
 PR
 PI NOBORU YAMAJI, KOICHI NISHIMURA, MIHO SASAMATA

PC C12N15/09,C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/10, PC
 C12N9/50,C1201/37,
 PC C12N15/00,C12N5/00
 CC
 FH Key Location/Qualifiers
 FT source 1..3618
 FT /organism="Homo sapiens"
 /db_xref="taxon:9606"
 ORIGIN
 BASE COUNT 1035 a 802 c 880 g 901 t
 FEATURES
 source 1..3618
 Location/Qualifiers
 /organism="Homo sapiens (human)"
 alignment_scores:
 Quality: 3624.00 Length: 1198
 Ratio: 4.081 Gaps: 17
 Percent Similarity: 74.124 Percent Identity: 55.593
 alignment block:
 US-09-938-330-20 x E58661 ..
 Align seg 1/1 to: E58661 from: 1 to: 3618
 37 AspTYGlyValThrValProCysSerThrAspPheArglyArgPheLe 53
 121 GAGTATGACCTGGTCTGACCTGACGCAATGCTAGAAAGAGCCTATCT 170
 53 uSerHisValValSerGlyProAlaAlaAlaSerAlaGlySerMetValV 70
 171 CTCCCATGCTCTTCT..... 186
 70 AlaSPThrProProThrLeuProArgHisSerSerHisLeuArg..... 84
 187ValAlaArgSerProLeuHisProGlyThrLeuTrpPr 98
 85ValAlaArgSerProLeuHisProGlyThrLeuTrpPr 98
 208 GCGAGGAGCAGTCTTCCACCCCT..... 231
 98 oGlyArgValGlyArgHisSerLeuTrpPheAsnValThrValPheGlyL 115
 232GAGCAGTTGTTCTTTAACATCAGCGCATTTGGAA 265
 115 ySGluLeuHisLeuArgLeuArgProAsnArgArgLeuValValProGlyL 131
 266 AAGATTTCATCTGGAGCTAAAGCCCAACACTAGTAGCTCCG 315
 132 SerSerValGluTrpGlnGlu..... 138
 316 GCTGTGTGGAGTGGCATGAGACATCTCTGTGCTGGCAATATTAACCGA 365
 139AspPheArgGluLeuPheA 145
 366 TCCCATTAACAACATGACACGAGAACTGCTACGTTAGCAATCCGGAATA 415
 145 rGlnProLeuArgGlnGlyCysValTrpThrGlyValAlaThrGlyMet 161
 416 CAGAGCTTTGGCAGACTAAGTCTATGTGTGATCGATCGAGCAATT 465
 162 ProGlyAlaAlaValAlaIleSerAsnGlyAspGlyLeuAlaGlyLeu 178
 466 CCAGAAACCTCTGTGGCATGAGCAACTGATGATGCTGCTGGCAATGAT 515
 178 eArgThrAspSerThrAspPhePheIleGluProLeuGluArgGlyGlnG 195
 516 AAAAGATGATTAATGACAGATATTTCATGTAACCCCTTGAAAGATGTAAC 565
 195 lngLysGluAlaSerGlyArgThrHisValValTrpArgArgGluAla 211
 566 AGATGGAGAAAGAAAGAGATTCATGTTGCTTACAAAGATCAGCT 615

212 ValGlnGlnGluTrpAlaGluProAspGlyAspLeuHis.....As 225
 616 GTAGAACAGAGCTCCCATGACATGTCACAAAGACTTCCATCAGAGATC 665
 225 nGluAlaPheGlyLeuGlyAspLeuProAsnLeuLeuGlyValGlyA 242
 666 GGACCTGGAAAGGCTTGTATGATGATGATGATGATGATGATGATGATG 715
 242 sPGLnLeuGlyAspThrGluArgLysArgArgHisAlaLysProGlySer 258
 716 AGCAGCTGATGATGAACATGAGACCCCGAGACACCGGAGAAACCAT 765
 259 TyrSerIleGluValLeuLeuValAlaAspAspSerValAlaArgPhe 275
 766 TACATATGAGAGTATGCTGTGGAGTGTGATGATGATGATGATGATGAT 815
 275 sGlyLysGluHisValGlnAsnTrpValLeuThrLeuMetAsnIleVal 292
 816 TGCCAAAGAGCAGTCACAAACTACCTCTGACCTTAATGAAATTTGTA 865
 292 sPGLuIleTrpHisAspGluSerLeuGlyValHisIleAsnIleAlaLeu 308
 866 ATGAATTTTACCATGATGATGATGATGATGATGATGATGATGATGATG 915
 309 ValArgLeuIleMetValGlyTrpArgGlnSerLeuSerLeuIleGluAr 325
 916 GTGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 965
 325 gGlyAsnProSerArgSerLeuGlnGlnValCysArgTrpAlaHisSerG 342
 966 GGGAAACCATTCACAAAGCTTGGAAATGTTGCTGGGGCGTCCCAAC 1015
 342 lngLArgGlnAspProSerHisAlaGlnHisHisAspHisValAlaPhe 358
 1016 AGCAAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 1065
 359 LeuThrArgGlnAspPheGlyProSer.....GlyTrpAlaProVa 372
 1066 TTAAACAGGCAAGACTTTGGACCTGCGAATGCAAGATATGCTCAGT 1115
 372 ThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisGlnAsp 389
 1116 CACCGGAGATGCTGATGATGATGATGATGATGATGATGATGATGATG 1165
 389 lPheSerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 405
 1166 GTTTTTCATCTGCTTTGTTGATGATGATGATGATGATGATGATGATG 1215
 406 MetGlnHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeu 422
 1216 ATGAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 1265
 422 ySerValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpS 439
 1266 AAGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1315
 439 eArgCysSerLysLeuGlnLeuSerArgTrpLeuProSerTrpAspCys 455
 1316 CCGCATGACAGGCTCAAGAACTGAAAGATATATCCATTCCTATGACTGT 1365
 456 LeuLeuAspAspProPheAspProAlaTrpProGlnProProGluLeuPr 472
 1366 CTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1415
 472 oGlyIleAsnTrpSerMetAspGluGlnCysArgPheAspPheGlySerG 489
 1416 TGGAAATCAATTTTCTATGATGATGATGATGATGATGATGATGATG 1465
 489 lYTrpGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeu 505
 1466 GCTATAAATGTGACCGGCTTCCGAACCTTTGACCCATGATTAACAGCTG 1515
 506 TrpCysSerHisProAspAsnProTrpPheCysLysThrLysGlyLpr 522

1105 oGlyProAspProGly.....ProThrSerLeuProProhe. 1117
 1118SerThrProGly.....SerProLeu 1124
 3278 ATTGACAGACCCCTGCAAGAGATGCTTTGAGTACAGATCTCTTCAGTG 3327
 1125 ProGlyProGlnAspProAlaAlaGluProProGlyLysProth 1141
 3328 GGAGGTCCAAATGACATATGCT...GCTTTCAGGCCAAACAGTAACCTGA 3374
 1141 rGlySerGluAspHisGlnHisGlyArgAlaThrGlnLeuProGlyAla 1158
 3375 TGTGCTCTAATTACCCGAC.....AGAGTGTCTCAGCAGCAGGAGGATA 3418
 1158 euAspThr.....SerSerProGlyThrGln 1166
 3419 AGACTGTGAGACTGTGACCGTACCATCTCCACCACCCACCAAG 3462
 seq_name: gb_pr:AB002364
 seq_documentation_block:
 LOCUS AB002364 5774 bp mRNA linear PRI 06-OCT-2001
 DEFINITION Human mRNA for KIAA0366 gene, partial cds.
 ACCESSION AB002364
 VERSION AB002364.1 GI:2224672
 KEYWORDS KIAA0366.
 SOURCE Homo sapiens male brain cDNA to mRNA, clone_11b:pbuescriptII SK plus clone:HH0124.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 5774)
 Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O., Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
 JOURNAL DNA Res. 4 (2), 141-150 (1997)
 MEDLINE 97345981
 REFERENCE 2 (bases 1 to 5774)
 AUTHORS Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
 TITLE Direct Submission
 JOURNAL Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
 FEATURES
 (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
 Location/Qualifiers
 1..5774
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HH0124"
 /sex="male"
 /tissue_type="brain"
 /clone_lib="pbuescriptII SK plus"
 1..3607
 /gene="KIAA0366"
 <1..3607
 /gene="KIAA0366"
 /codon_start=2
 /db_xref="GI:2224672"
 /translation="STMIAALVEVTSADGAGNEENVQIDLPTRRYRELVTPV
 STNLEGRSHSTLSASHKRSARVSNBPDLFPNTAFGDFHLKRPNTOLVAGA
 VVEWHTSLVPGNTIDPILNNHQPQSAVYLRTEPIQTNCAVYGDIDVDPQSVAISN
 CDGLAGMIKSDNEYETLEPGRKOMEKGRHVVYKRSAYVADIDSKSPHVS
 LLEGIDLDGTYVNIHOOLNETMRBRHAGENVYNEVLGVDSVYFRGKHVONY
 LTETMINVETVHDESHGVAHNVYLVMTILGAKTSILIERGNPSRSLENCMAASQ
 QORSDLNHSEHDAIFLTRQDFGPAGMOGYAVNTGMCHEPVSCTLNHEDGSSAFVY
 AHETGVHVGMEHDQGRGDETMGSMVAPLVQAAFPNRYHMRSGSGDLKRYISYD
 CLDDPDHDMPKLPELPGINGYSMDEGRDFPDVGVKMGCAFPETDPCOLMCSPDN
 PYFCKTKKGPPLDTECAAGKWCYKKGHCMMKNANQCKQDGNMCSWTKPSCSRCTGTC

VFRTQCNPNMPEINGGDCPGVNEFYOLCNTSECKHFEDEFRQAQCCORNSHREYON
 TKHMLPJEHPKPKRCHLYCOSKETGADYOLVHKLVDHGCYSKDPYSICVRECYK
 VCGDKIEGSKNKGKCGVCGGDNHSCRTKVGFTFTPKRIGYLIKMFDPGARHVILLO
 EDEASPHILAIKNQATGHIILNGKGEARKSRFTDILGYEMOYNTIEDLIESLHTGPH
 DPVYVLIIPQENDTRSLTKYKTHIHEDSPPTINSNVYIOBELDFEALAKMSQVSRP
 CGGGRFYTTGCRKRSKDNKMFVRSFCENAKKPKPIRMNCNIOECTHPLMAVEWHCT
 KTGSSGYOLRTVRCLOPLDGTNRSVHSHKVCMDRPSRRPCNRPVCPAOMKTPWS
 ECSVTCGEGTEVROYLCRAGDHDEKPEVYACOLPPCNDPCLGDKSIFCOMEVLA
 RYCSILPGYKRLCCESCSKRSSTLPPPYLLKAEETHDVIYSNPSDLPRSLVMPSTLVY
 HSETPAKKMSLSSISVSGPNAYAFRPNRSPDGNANRORSQAQAGSKTYRVLTPVPS
 PPKRYHLSASOMAAFPFASDSIGASSQARSKRKDKITIDNRPRSTLRR"
 BASE COUNT 1766 a 1129 c 1258 g 1621 t
 ORIGIN
 alignment_scores:
 Quality: 3615.50 Length: 1274
 Ratio: 3.956 Gaps: 22
 Percent Similarity: 71.743 Percent Identity: 53.373
 alignment_block:
 US-09-938-330-20 x AB002364 ..
 Align seg 1/1 to: AB002364 from: 1 to: 5774
 37 AspyrGlyValThrValProGlySerThrAspPheArgGlyArgPheLe 53
 110 GAGTATGAGCTGTGACTGCTGACGACCAAAATCTGAAAGAGCGTATCT 159
 53 userHisValSerGlyProAlaAlaAlaSerAlaGlySerMetValV 70
 160 CTCCCATACCTCTTCT..... 175
 70 alaThrProProThrLeuProArgHisSerThrLeuArg..... 84
 176ValAlaArgSerProLeuHisProGlyGlyThrLeuTrp 98
 85ValAlaArgSerProLeuHisProGlyGlyThrLeuTrp 98
 197 GCGAGGACGCTGTTCACACCT..... 220
 98 oGlyArgValGlyArgHisSerLeuTyrrPheAsnValThrValPheGlyL 115
 221GAGCAGTTGTTCTTTAACATCAGCGCATTTGGGA 254
 115 ysgIuLeuHisLeuArgLeuArgProAsnArgArgLeuValValProGly 131
 255 AAGATTTTCATCTGCGACTAAAGCCCAACACTCACTAGTACTCTCTGGG 304
 132 SerSerValGluTrpGlnGlu..... 138
 305 GCTTGTGTGAGTGGCATGAGACATCTCTGCTCGGGAATAATACCA 354
 139AspPheArgGluLeuPheA 145
 355 TCCCATTAACACCATCAACAGAGAGTCTAGTATAGATCCGGAATA 404
 145 rginProLeuArgGlnGluGlyValTyrrThrGlyGlyValThrGlyMet 161
 405 CAGAGCGCTTTCAGACTACTGTGCTTATGTTGGTACATCTGTGACATT 454
 162 ProGlyAlaAlaValAlaIleSerAsnGlyAspGlyLeuAlaGlyLeu 178
 455 CCAGGAACCTCTGTTCCATCAGCACTGTATGCTGCTCGGGAATAT 504
 178 eaGlyThrAspSerThrAspPhePheIleGluProLeuGlnArgGlyGln 195
 505 AAAAAGTGAATGAAGACTATTTCATTCAGCCCTTGAAGAGAGTAAAC 554
 195 InGluysGluAlaSerGlyArgThrHisValValTyrrArgArgGluAla 211
 555 AGATGAGAGAGAAAGAGAGATTCAATTGTTGTTACAGAGATCAGCT 604

[illegible]

2402 GTTATTGTTTGGATTATA...CCTCAAGAAAATGATACCCGCTTAGCCT 2448
 822 UAlATyLysTyValIleHisGlnspleuProleuIleGlySerA 839
 2449 GACATTAATACATCATCATGAAGACTCTGACTACATCAACACACA 2498
 839 snAsnValleuGlnMetAspThrTyrluTrpAlaLeuLysSer 855
 2499 ACAATGTCATCCAGAGAAATTAAGTACTTTGAGGGGCTTTGAGAGAC 2548
 856 TrpAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTy 872
 2549 TGGTCTCAGAGTTCCAAACCCCTGCTGGAGGTTTCCAGTACATAAATA 2598
 872 rGlyCysArgArgArgAspHisHisMetValGlnArgHisLeuCysA 889
 2599 TGGATCCCTAGAGAAAGTGAATTAATAATGTTCCATCGCAGCTCTGTG 2648
 889 sPhisLysArgProLysProIleArgArgArgCysAsnGlnHisPro 905
 2649 AGGCCAACAACAAAGCCGAACCTATTAGACGAATGCAATATTCAAGAG 2698
 906 CysSerGlnProValTrpValThrGlnGlnTrpGlyAlaCysSerArgSe 922
 2699 TGTACATTCACCTCTGGGTACAGAGAAATGGAACACATGACACCAAAAC 2748
 922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuProL 939
 2749 CTGTGAATCTCTGGGTATACACTTCGCACATGACGCTGCTTCAACCCAC 2798
 939 euserAsnGlyThrHisLysValMetProAlaLysAlaCysAlaGlyAsp 955
 2799 TCCGTGATGCGACCAACCGCTCTGTGCACAGCAATACTGCATGAGGTAC 2848
 956 ArgProGlnAlaArgArgProCysLeuArgValProCysProAlaGlnTr 972
 2849 CGTCCCGAAGAGCCGCCGCCCTGTACAGAGTGCCTGCTGACACAGTG 2898
 972 pArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyGlnGlyLeG 989
 2899 GAAACAGAGACCCCTGAGTGAAGTTTCATGACCTCGCGGTAAAGAACGC 2948
 989 lngLArgGlnValValCysArgTrpAsnAlaAsnSerLeuGlyHisCys 1005
 2949 AGGTAGAGGAGGTCTCTGACAGGCGTGGC.....GACCACCTGT 2986
 1006 GlnGlyAspArgProAspThrValGlnValCysSerLeuProAlaCysG 1022
 2987 GATGGTGAAGACCTGAGTGCCTCAGACCTGCTCAACTGCTCTTGTAA 3036
 1022 yGlyAsnHisGlnAsnSerThrValArgAlaAspValTrpGlnLeuGlyT 1039
 3037 TGAAT..... 3040
 1039 hrProGlnGlyGlnTrpValProGlnSerGlyProLeuHisProIleasn 1055
 3040 3040
 1056 LysIleSerSerMetCysAlaAlaGlnProCysThrGlyAspArgSerVa 1072
 3041GAACCATGTTGGGAGAACAGATCCAT 3066
 1072 lPheCysGlnMetGlnValLeuAspArgTyCysSerIleProGlyTythr 1089
 3067 ATTCTGTCAATGGAAGTGTGGCAGATACTGCTCATACCAAGGTTATA 3116
 1089 lAsArgLeuCysCysValSerCysIleLysLysAlaSerGly..... 1102
 3117 ACAAGTTATGTTGTAGTCTCGACAGCAAGCGACGTACACCCGACCAACA 3166
 1103ProAsnPr 1105
 3167 CCATACCTTTAGAAAGCTGCTGAACACTCATGATGATCTCATCTTAACCC 3216

1105 ogGlyProAspProGly.....ProHisSerLeuProProphe. 1117
 3217 TASTGACCTCCCTAGATCTCTAGTATGCTGCTACATCTTTGGTCTTATC 3266
 1118SerThrProGly.....SerProLeu 1124
 3267 ATTCAAGAGACCCCTGCMAAGAGATGCTTTGATGATGACATCTCTTCACTG 3316
 1125 ProGlyProGlnAspProAlaAspAlaIleGlnProProGlyLysProTh 1141
 3317 GAGAGTCCAAATGCAATATCT...GCTTTCAGGCCCAAAACAGTAAACCTGA 3363
 1141 rGlySerGlnAspHisGlnHisGlyArgAlaThrGlnLeuProGlyAlaL 1158
 3364 TGGTCTCAATTTTACGCCAG.....AGGAGTCTCAGCAACAGAGAGATA 3407
 1158 eusAspThr.....SerSerProGlyThrGlnHisPro 1168
 3408 AGACTGTGAGACTGTGACCCGTACCATCTCCCAACCAAGAGAGGTC 3457
 1169PheAlaProG 1172
 3458 CACCTCAGTTCAAGCTTCAACAAATGCGTCTGCTCTTCTTTCAGACCCAG 3507
 1172 uThrProIleProGlyAlaSerTrpSerIleSerProThrThrProGlyG 1189
 3508 TAAATCAATAGTGTCTTCTTCAGCAGCAAGAACCTCAAAAGATGAA 3557
 1189 lYeu.....ProTrpGlyTrpThrGlnThrProThr... 1199
 3558 AGATCAATTGACACAGACAGCTCG.....AlaAspLysGlnProG 1209
 1200ProValPro.....GlnAspLysGlnProG 1209
 3599 GAAAGATGAGAAAGTGAACCAAAAGCTGACAAACAGAGAGAAACCTGG 3648
 1209 yGlnAspLeuArgHisProGly 1216
 3649 ACAACCTCTCTCTTCCATGCT 3670

seq_name: gb.com:BTCPINPcn
 seq_documentation_block: 4580 bp mRNA linear MAM 01-MAR-1997
 LOCUS BTCPINPcn
 DEFINITION B.taurus mRNA for procollagen I N-proteinase.
 ACCESSION X96389
 VERSION X96389.1 GI:1865715
 KEYWORDS PC I-NP gene; procollagen I N-proteinase.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 4580)
 Colige A., Nusgens B. V. and Lapiere C.M.
 Cloning of the cDNA of the bovine procollagen I N-proteinase
 2 (bases 1 to 4580)
 Colige A.C.
 Direct Submission
 Submitted (29-FEB-1996) A.C. Colige, University of Liege,
 Laboratory of Connective Tissue Biology, Tour de Pathologie, B23/3
 - CHU Sart-Tilman, B-4000 Sart-Tilman, BELGIUM
 Location/Qualifiers
 1..4580
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /tissue_type="skin"
 /dev_stage="foetal"
 138..3755
 /gene="PC I-NP"
 CDS
 138..3755


```

429 ValGlnAlaIaIaPheHisArgPheHisIleTrpSerArgCysSerIleLeuG1 445
1434 GTGACAGACACCTTCATCGCTCCACTGCTCCCGTGGACGACACAGAGA 1483
445 uIeuSerArgTyrLeuProSerTyrAspCysIleuLeuAspPheProPhea 462
1484 GCTTACCGCTTACCTGCTCTATGACGCTGGCGGATGACCCCTTCA 1533
462 sPProIaIaTrpProGlnProGluLeuProGluIleAsnTrpSerMet 478
1534 CCCAGACACTGGCGGCTGGCTGCCAGCTCCGGGCTGGACACTCCATG 1583
479 AspGluGlnCysArgPheAspPheGlySerGlyTyrGlnThrCysLeuA1 495
1584 AACGACAGACTGGCGCTTGCATTTGGCGCTGGTACATGATGTGTACCGC 1633
495 aPheArgTrpPheGluProCysGlnLeuTrpCysSerHisProAspA 512
1634 GTTCCGACCTTGGACCCGCTGCAACACGCTGTGTGACGACCCCTGACA 1683
512 snProTyrPheCysLysThrIleLysGlyProProLeuAspGlyThrGlu 528
1684 ACCCTACTTTTGAAGAACAAAGAGGGGCCACCTCGATGGACCATG 1733
529 CysAlaProGlyLysTrpCysPheLysGlyHisCysIleTrpLysSerP 545
1734 TGTGGCCCTGGCAAGCAGCTGTAAAGACACTGCATCTGGCGTACACCC 1783
545 oGluGlnThrTyrGlyLysAspGlyLysTrpSerSerTrpThrLysPheG 562
1784 TATAC...ATTTCAAACAGAGAGGCACTGGGCGCTGTGAGTCCCTTGC 1830
562 LysCysSerArgSerCysGlyGlyValArgSerArgSerArgSer 578
1831 GCTCCTGCTGGCTACGTACGCGACAGGTGTGAGTTACAGACCCGCTCAG 1880
579 CysAsnAsnProSerProAlaTyrGlyLysArgProCysLeuGlyProme 595
1881 TCCGACAAACCCACACCCAGCCAAATGGGGCGCCACATGCTGGCGCTCGC 1930
595 rPheGluTyrGlnValCysAsnSerGluLysCysProGlyThrTyrLysL 612
1931 CTACGATTTCCAGCTCTCTCAACTGCGAGACTGCTGACGCGCTGGCCG 1980
612 sPPhaIaIaGlnGlnCysAlaIaLysArgAsnSerTyrTyrValHisGln 628
1981 ACTTCCGAGAGAGAGTGGCGGACGTGCTGCTTCCAGCATGCT 2030
629 AsnAlaLysHisSerTrpValProTyrGluProAspAspAlaGlnLys 645
2031 GACGCTCAACACCACTGGCTGCCACGACGACCCGGAGCCCAAGGACGC 2080
645 sCysGluLeuIleCysGlnSerAlaAspThrGlyAspAlaValPheMetH 662
2081 GTGTGATCTCTACTGTGAGTCCAGAGACCCGGGAGGTGTGTCCATGTA 2130
662 snGlnValValHisAspGlyThrArgCysSerTyrArgAspProTyrSer 678
2131 AGCGTATGGTCCATGACGGGACAGCTGTCTTCAAGGACGCTTACAG 2180
679 ValCysAlaArgGlyLysCysValProValGlyCysAspLysGluValG1 695
2181 CTCTGGCTGGCTGGGACTGCGAGAGAGTGGCTGTGACGGGTGTATCGG 2230
695 ySerMetLysAlaAspAspLysCysGlyValCysGlyGlyAspAspSerH 712
2231 CTCACAGCAGCAGGAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2280
712 LysCysArgThrValLysGlyThrLeuGlyLysAlaSerLysGlnAlaGly 728
2281 ACTGCAAGGTGTGTAAAGGCAAGTGTGTGTGTGTGTGTGTGTGTGTGT 2330

729 AlaLeuLysLeuValGlnIleProAlaGlyAlaArgHisIleGlnIleG1 745
2331 TACATAGAGATGTTTATAGATCCCGGACGAGGACCAACACTGCTAATCCA 2380
745 uAlaLeuGluLysSerProHisArgSerValValLysAsnGlnValThrG 762
2381 GGAAGCAGACACACACGACCATCACCTGGCCCTCAAAAACCTGGAGACAG 2430
762 LysPheIleIleuAsnProLysGlyLys...GluAlaThrSerArgThr 777
2431 GCAAGTTTCAATTTTAAATGAGGAATGACGTGATCCCACTCCAAAGACC 2480
778 PheThrAlaMetGlyLeuGluThrProGluAspAlaValGluAspAlaLysG1 794
2481 TTCATGCGCATGGGCTGGAGTGGAGTACCGGGATGGAGGACGCGCGGA 2530
794 uSerLeuLysThrSerGlyProLeuProGluAlaIleAlaIleLeuAla1 811
2531 GACGCTGACAGACCATGGGCGCTCCAGGCAACATCATCTGCTGTGTCA 2580
811 euProProThrGluGlyLysProArgSerSerLeuAlaTyrLysTyrVal 827
2581 TCCCA...GAGGGGACGCGCCGACATCTCATCTGACCTTACAAATACATG 2624
828 IleHisGluAspLeuLeuProLeuIleGlySerAsnAsnValLeuLeuG1 844
2625 ATCCATGAGAGACTCCCTTAAT...GTGGATACAAACACGCTCGGAAGA 2671
844 uGluMetAspThrTyrGluThrAlaLeuLysSerThrAlaProCysSerL 861
2672 CGACTCTGTGGCTATGATGGGCGCTGAAAGATGGTGGCTGGCTGCTCA 2721
861 yAlaCysGlyLysGlyLysIleGlnPheThrLysTyrGlyCysArgAlaArg 877
2722 ACCCTGCGGTGAGGAGTCCCAATTCACCAAGTATGCTGGCGCGGAGG 2771
878 ArgAspHisIleMetValGlnArgHisIleuCysAspHisLysLysArgP 894
2772 CTGACACCAACAGATGTGTGACCGAGGCTTGTGACATCCGCTCAAAAGCC 2821
894 oLysProIleArgArgArgCysAsnGlnHisProCysSerGlnProValT 911
2822 CAAGCCATCCCGCGGACCTGCAACCCACAGAGAGTGTCCACGACCGGTG 2871
911 rPValThrGluGluThrProLysAlaCysSerArgSerCysGlyLysLeuGly 927
2872 GGGTACGCGGTGAGTGGAGCGGTGACCGGAGAGCTGTGGCGGACAGGC 2921
928 ValGlnThrArgGlyIleGlnCysLeuLeuProLeuSerAsnGlyThrH1 944
2922 ATGCAAGTGTGCTGTGTGGCGTGTGTACCTGTGTGACAAACACACAC 2971
944 sLysValMetProAlaLysAlaCysAlaGlyAspArgProGluAlaArgA 961
2972 CCGCTCGGTGACACCAAGCACTGCAATGACGCTGACCCGAGGCGCGCC 3021
961 rProCysLeuArgValProCysProAlaGlnThrPheGluGlyAlaTrp 977
3022 GGGCTGCAACCGGAGACTGTGCTGGCGGTGTGCGGAGGAGCGCGCTG 3071
978 SerGlnCysSerAlaThrCysGlyGlyGlyIleGlnGlnArgGlnValA 994
3072 TCCAGTGTCTACATTAACCTGTGAAGAGGACCCAGCAAGCGGCAATGCT 3121
994 LysArgThrAsnAlaAsnSerLeuGlyHisCysGlyLysAspArgProA 1011
3122 CTGCGCAACTGCGGACGACAGTTCGAGGTGTGCGGAGGAGCGCGCTG 3171
1011 sPThrValGlnValCysSerLeuProAlaCysGlyLysLysHisGlnLys 1027
3172 AGAGCGCAAGATCTGAGGCTTGGCGCTGTGCCCGCAACACCTGTGAC 3221
1028 SerThrValArgAlaAspValTrpGluLeuGlyThrProGluGlyLysL 1044

```


175 GCGGTGCCCGCTGCGCACTGACGCCAGGCCGCTGGTGTCCAGCTGG 224
 57 ISeGlyProAlaIaIaIaSerAlaGlySerMetValAlaAspThrProp 74
 225 GTCCGGCA...GCTAGCTCCAGAGAGGGGTACAGAGCCCGAGGGCGGCC 271
 74 roThrLeuProAlaGlnHisSerSerHisLeuArgValAlaArgSerProleu 90
 272 CG.....GTCCGGAGCCCGAGAC 288
 91 HisProGlyGlyThrLeuThrProGlyArgValGlyArgHisSerLeuTy 107
 289 TTCCCGCGAGGCAAC.....GAGGAGAGCCCGGACATCAC...CTCTT 329
 107 rPheAsnValThrValPheGlyLysGlnLeuHisLeuArgLeuArgPro 124
 330 CTACAAATGTCACGCTCTTGGCCGAGACCTGCACCTGGCGTGGCCCA 379
 124 snArgArgLeuValValProGlySerSerValGluTrpGlnLysPhe 140
 380 ACGCCGCGCTCGTGGCGCCGCGGACCTATGGAGTGGCAGGGCGAGAG 429
 141 ArgGlnLeuPheArgGlnProLeuArgGlnGlyValTyThrGlyG 157
 430 GGCACCAACCGCGCTGGAGCCCTCGGAGCTGTCTCTACGTGGAGA 479
 157 yValThrGlyMetProGlyAla...AlaValAlaIleSerAsnGlySasp 173
 480 CGTGGCGCGGCTATGCCAGAGCCCTCTCTGTGGCGCTCAGCAACTCGATG 529
 173 LysLeuAlaGlyLeuLleArgThrAspSerThrAspPheIleGluPro 189
 530 GCGTGGCTGTCTGATCCGATGAGGAGGAGGAGGAGCTCTTCATGCAACC 579
 190 LeuGlnArgGlyGlnGlnGlnLysGlnLys...SerGlyArgThrHisVa 205
 580 TTGGAAAGAGGGCTGGCGGCGGAGAGGCTGAGCAAGGCCGTGTGCATGT 629
 205 LValTyThrArgGlnAlaValGlnGlnGlnLysTrpAlaGluProAspGly 221
 630 GGTGATCCCGCGCCACCCACCTCCCTCTCGGGGGCGCACAGGCC 679
 222 ..AspLeuHisAsnGlnAlaPheGlyLeuGlyAspLeuProAsnLeu 237
 680 TGGACACAGAGGGCTCCCTGGACAGCTGAGACGCTCAGCGCGCCCTG 729
 238 GlyLeuValGlyAspGlnLeuGlyAspThrGlnArgLys...ArgArgHis 253
 730 GCGTCTCTAGAGGAGCAGCGCAACAGCTCGAGGGGAGGAGGACCGACGA 779
 253 sAlaLysProGlySerTyrrSerIleGlnValLeuLeuValAlaAspSps 270
 780 TCTGTGACAGCATGTGACTACAACTGAGAGTCTCTGTGGCGTGGATGACT 829
 270 eValValArgPheHisGlyLysGlnHisValGlnAsnTyrrValLeuThr 286
 830 CTGTGTGTCAGTTCACGAGGAGGAGGAGCTGACAGAACTACTCTTACA 879
 287 LeuMetAsnIleValAspGluLleTyrrHisAspLysLeuGlyValHis 303
 880 CTCATGAAACATTGTCAATGAATCTACCATGACGAGCTTGGGTGCCCA 929
 303 sIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrrArgGlnSer 320
 930 CATCAAGCTGTCTCTGTGGCGGATCATCTCTCTAGCTATGGAAGATCCA 979
 320 euSerLeuIleGluArgGlyAsnProSerArgSerLeuGlnGlnValCys 336
 980 TGAGCTCATGAGATCGGAAACCTCTCTCAGAGCCTTGGAATGTGTGC 1029
 337 ArgTrpAlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHis 353
 1030 CGGTGGCGCTTACTCTCAGAGAGCAGACAGGGCCACGATGAATACCA 1079
 353 sAspHisValValPheLeuThrArgGlnAspPheGlyProSer..... 367
 1080 CGATCAGCATCTTCTCTCAGAGGAGGAGCTTGGCGCTTCCGGGATTC 1129
 368 ..GlyTyrrAlaProValThrGlyMetCysHisProLeuArgSerCysAla 383
 1130 AAGGTAATGCTCTGTACACCGGATGTGCACTCCGCTCCGAGCTGAC 1179
 384 LeuAsnHisGlnAspGlyPheSerSerAlaPheValIleAlaHisGln 400
 1180 CTGACCATAGAGACGGCTTCTCTCAGCGCTTGTGGTGGCCATGAGAC 1229
 400 rGlyHisValLeuGlyMetGlnHisAspGlyGlnGlyAsnGlyCysAla 417
 1230 TGGCCAGCTGTGGCATGAGACAGCAGGAGGAGCAACCTGTGGCG 1279
 417 spGluThrSerLeuGlySerValMetAlaProLeuValGlnAlaPhe 433
 1280 ACAGAGGTGGCGCTGGGACATCATGTGGCGCCCTGTGAGGCCCTTC 1329
 434 HisArgPheHisTrpSerArgCysSerLysLeuGlnLeuSerArgTyrr 450
 1330 CACCGCTTCCACAGTGTCCGCTGACAGCCAGCAGAGCTGAGCCGCTACT 1379
 450 uproSerTyrrAspCysLeuLeuAspAspProPheAspProAlaTrp 467
 1380 GCATCTCTATGACTGTCTGTGATGACCCCTTCCCGACAGCATGGCGG 1429
 467 InProGluLeuProGlyIleAsnTyrrSerMetAspGlnGlnCysArg 483
 1430 CGGTGGCCCGAGCTCCCGGAGCTGCACTACTCTCATGAGAGCAATGCCG 1479
 484 PheAspPheGlySerGlyTyrrGlnThrCysLeuAlaPheArgThrPhe 500
 1480 TTGTACTGTGGCTGGCTGATCATGTGACAGGGGCTTCGAGACTTTGA 1529
 500 uproCysLysGlnLeuTrpCysSerHisProAspAsnProTyrrPheCys 517
 1530 CCCTGTCAAGACAGCTGTGGTGGAGCCATCTGACAAACCTTACTTTGCA 1579
 517 yThrLysLysGlyProProLeuAspGlyThrGlnCysAlaProGlyLys 533
 1580 AGACAAAGAGGGGCGCCCTTGGACGGAGCTATGTGTGACCTGGGAG 1629
 534 TrpCysPheLysGlyHisCysLleTrpLysSerProGlnInThrTyrr 550
 1630 CATGTCTTAAAGGACACTGTGATCTGGCTGACACCTGAC...ATCTCAA 1676
 550 yGlnAspGlyGlyTrpSerSerTrpThrLysPheGlySerCysSerArg 567
 1677 ACGGAGCGGAGCTGGGGCGCTGTGAGTCCGCTTGGCTCTGCTCACGTA 1726
 567 eTyrrGlyGlyValArgSerArgSerArgSerCysAsnAsnProSer 583
 1727 CCGTGTGACAGCGCGGTGAAGTTTCAGAGCCCGCAGTGTGACAAACCCAC 1776
 584 ProAlaTyrrGlyArgProCysLeuGlyPrometPheGlnTyrrGlnVa 600
 1777 CGGGCAACGGGGGCGGACCTGCTCGGGCTTCCCTGACACATTCACAG 1826
 600 LysAsnSerGlnGlyCysProGlyThrTyrrGlnAspPheArgAlaGln 617
 1827 CTGACGCGCGCAGGACTGCCCCGACTCCCTGCGTACATTCAGAGAG 1876
 617 LncysAlaLysArgAsnSerTyrrTyrrValHisGlnAsnAlaLysHis 633
 1877 AGTGGCCGCAATGGAGCTTACTTCAGACGCGCAGCGCCAGCACAC 1926
 634 TrpValProTyrrGluProAspAspAlaGlnLysCysGlnLeuLeu 650
 1927 TGGCTGGCCCGCAGAGCACCGGAGTCCCAAGAGAGATGCCACTGTACTG 1976

650 sgnlserAlaasprthrGlyaspValValPheMetasnglValValHisA 667
 1977 CGAGTCCAGGAGACCGGGGAGGTGTGTCCATGAGGCGCATGTGTGATG 2026
 667 spgIYThrArgCysSerThrArgAspProLYrSerValCysAlaArgGly 683
 2027 ATGGGACGCGGTCTCTACAGAGACGCTTCAGCTGTGTGTGTGCGGG 2076
 684 GlucysValProValGlyCysAspLysGlyValGlySerMetLysAlaAs 700
 2077 GACTGCAGAGAGGTGGCTGTGAGCTGTGATCGCTCCAGCAAGCAGAGA 2126
 700 pAspLysCysGlyValCysGlyGlyAspAsnSerHisCysArgThrValL 717
 2127 AGACAAGTGTGGCTGTGCGGAGGACACAGCCATGCAAGGGTCA 2176
 717 YsgIYThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuVal 733
 2177 AGGGACAGCTTCACAGCTTCACCCAGAGCATGTTACATCAAGATGTTT 2226
 734 GlnIleProAlaGlyAlaArgHisIleGlnIleGlyAlaLeuGlyLysSe 750
 2227 GAGATTCCTCGACAGACCCAGACCTGCTCATTCAGAGGTAGAGCCGAC 2276
 750 rPhnAsArgSerValValLysAsnGlnValThrGlySerPheIleLeuA 767
 2277 CAGCCACCATGTGGCGGTCAAGACCTGGAGACAGGCAAGTTCATCTTA 2326
 767 snProLysGlyLys...GluAlaThrSerArgThrPheThrAlaMetGly 782
 2327 ATGAGAGAGATGACGTGATGCCAGATCCAAACCTCATTCATGCCATGGGC 2376
 783 LeuGluThrGlyAspAlaValGlyLysAlaLysGlySerLeuLysThrSe 799
 2377 GTGGAGTGGAGTACAGAGACGAGCGGCGGAGCGCTGCACAGCCAT 2426
 799 rGlyProLeuProGlyAlaIleAlaIleLeuAlaLeuProProThrGluG 816
 2427 GGGCCCCCTCCAGGACCATCACCCTTGTGTGTCATCCG.....GTGG 2470
 816 YgLYrProArgSerSerLeuAlaTYrLysTYrValIleHisGlyLysLeu 832
 2471 GAGACACCGGGGTCTCAGTACGATCAATATCATGATGAGGACTCA 2520
 833 LeuProLeuIleGlySerAsnAsnValLeuGluGluMetAspThrTY 849
 2521 CTGAAT...GTGATGACAAACAGTCTCGAGAGAGGACTGTGTGTCTA 2567
 849 rGluThrAlaLeuLysSerThrAlaProCysSerLysAlaCysGlyGlyG 866
 2568 CGAGTGGGCGCTGAGAGAGTGTCTCGTGTCCAGGCCCTGTGGCGGAG 2617
 866 YrIleGlnPheThrLysTYrGlyCysArgArgArgArgAspHisMet 882
 2618 GTGCCAGTTCACCAAGTATGCTGCCGCCGAGGCTGAGCAACAAGATG 2667
 883 ValGlnArgHisLeuCysAspHisLysLysArgProLysProIleArgAr 899
 2668 GTACACCGT 2717
 2718 AGGGTGCAGACCCACAGGATGTCTCCAGAGCATGTGTGTGTGTGTGTGT 2767
 916 rpeGlyAlaCysSerArgSerCysGlyLysLeuGlyValGlnThrArgGly 932
 2768 GGGAGCCATGTAGCCAGACCTGTGGGAGACAGCATGCGAGGTGCTCC 2817
 933 ILGlnIleLeuLeuProLeuSerAsnGlyThrHisLysValMetProAl 949
 2818 GTGGCGTGCATTCAGCGCGTACACAGACACACACCGCGCTGTGTGTGACG 2867
 949 aLYsAlaCysAlaGlyAspArgProGlyAlaIleArgArgProCysLeuArgV 966

2868 CAAGCACTGCATAGACCGCCCGGAGAGCCGCGGCGCTGCACCGCG 2917
 966 aProCysProAlaGlnThrArgLeuGlyAlaIlePheSerLysCysAla 982
 2918 AGCTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2967
 983 ThrCysGlyGlyGlyIleGlnArgGlnValValCysArgThrAsnAl 999
 2968 ACCTGTGGCAAGGACCCAGAGAGCGCCAGTCCCTGCGGACCGCGGA 3017
 999 aAsnSerLeuGlyHisCysGlyLysAspArgProAspThrValGlnValC 1016
 3018 CGACAGCTTGCATCTGCCAGAGAGAGCGCTCTGAGACAGGAGACCT 3067
 1016 YsSerLeuProAlaCysGlyGlyLysAsnHisGlnAsnSerThrValAla 1032
 3068 GCAAGCTTGGCGCTGTGCGGAGACATCTCAGATCCCTCCAGAAAGAGC 3117
 1033 AspValThrProLeuGlyThrProGlyGlyGlnThrValProGlnSerG 1049
 3118 TACGTA.....GTTCAGTGGCTGTGCCCGCGGA 3146
 1049 YrProLeuHisProIleAsnLysIleSerMetCysAlaIleGluProC 1066
 3147 CCCGACCTGCCCATCCGGAAGATCTGTCA.....AAGGCCACT 3187
 1066 YsrThrGlyAspArgSerValPheCysGlnMetGluValLeuAspArgTYr 1082
 3188 GCGAAGCGGACAGTCAATTTCTGTAGATGAGAGTCTGTGTCCGCTAT 3237
 1083 CysSerIleProGlyTYrHisArgLeuCysCysValSerCys..... 1096
 3238 TGCTCCATCCAGGCTTACAAAGCTGTCTGCAAGCTCTGTAACTGTGA 3287
 1097IleLysLysAlaSerGlyProAspProGlyPro.. 1107
 3288 CAAACACCTCACCAACGTGGAGGCGAGATGAGCCCGCTGGGAGAGC 3337
 1108AspProGlyProThrSerLeuProProPheSerThrPro 1120
 3338 ACAACGACATTTGACGTGTTCATGCTTACCTCCAGAGGCCACATGTAGC 3387
 1121 GlySerProLeuProGlyProGlnAspProAlaAspAlaIleGluProP 1137
 3388 ATGAGAGTGGCGCTCATCACCAAGCACCCCTGAGAGTCCCTGCAATGC 3437
 1137 oGlyLysProThrGlySerGlyAspHisGlnHisGlyAlaArgAlaThrGlnL 1154
 3438 CTCAGGACCAATGCCACAGAGATCACCCAGAAACCAATGCCGTAGATG 3487
 1154 euProGlyAlaLeuAspThrSerSerProGlyThrGlnHisPro...Phe 1169
 3488 AACCTTACAAATATCATGCGCTGGAGATGATGATGCCAGACCCAACTA 3537
 1170 AlaProGluThrProIlePro 1176
 3538 ATCCCTGACGACGAGGCC 3558
 seq_name: gb_pat:E55265
 seq_documentation_block: 5061 bp DNA linear PAT 31-JAN-2002
 LOCUS E55265 Novel metalloprotease and gene of the same.
 DEFINITION Novel metalloprotease and gene of the same.
 ACCESSION E55265
 VERSION E55265.1 GI:18629778
 KEYWORDS JP 2001008687-A/1.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 5061)
 AUTHORS Yamaji,N., Nishimura,K. and Sasamata,M.


```

435 gphenistpserargcysserlyleuileuSerArgTyrleuPro. 451
    ::::::::::::::::::::
1275 cctacacctgctccgctgacgcccagatmtacacaggttcttgacc 1324
452 ..SetIlyAspCysleuAspAspProPheAspProAla..... 464
1325 gtgggtgggacctgtgctggacgacctt.....cctggccagacatt 1368
465 .....TtrProGlnProProGluLeuProGlyIleAsnTyrSerMetAs 479
1369 atcgacttccctcggtgcca.....cctggcgtccttatgatgttaag 1412
479 pgluglucysargpheasppheglyserlytyrlnthrcysleuAlar 496
1413 ccacagctccgcttcacagtagcgg.....gcct 1441
496 hearghrhe.....GluProCysLysGlnleuTrrCys 507
1442 accttccttctgacgacatgatattctgcccacacacttgctgc 1491
508 serhisrproasparnprotyrphcyslysthrlysglyproProle 524
1492 tctgtggggacac.....tctacatccacagctgacagccgt 1532
524 uasrglythrlycysAlaProGlylysttrpysrphelysglyhisCys. 540
1533 gcacggcaccggtgtgggagatagtgcttcagtgggagctgcg 1582
541 .....IletrpLysserProGlnlnthrclyglnaspglylytrp 555
1583 tacccgggcttccgcccagagccgtg.....gatgtgctgcg 1623
556 SerSertrpThrlysrphelyserCysSerArgserCysglylytyla 572
1624 tctggctgagcgcctgctcattgctcagcagctgctgcagtcggcgt 1673
572 IargSerArgSerArgserCysAsnAsnProserProAlaTyrGlyIYA 589
1674 acacagcggcggagcagcagtcacacagcctaccccaattacaaagcca 1723
589 rgrProCysleuglyPrometrheglutyrGlnValCysAsnSerGluGlu 605
1724 gatcactgtgggtgagcgcacagcgtccgcttcacacacacagcgc 1773
606 CysProGlythrtyrGluAspPheArgAlaGlnlnCysAlaIlystAs 622
1774 tgccttcctggccacccctccttcctgcacagctccagtcagccacttga 1823
622 nserTyrTyrValHisGlnAsnAlaIlyHisSerTrrValProTyrGluP 639
1824 cggT...AtgctTACAAGGGCCAGTCGACACATGGGTGCGGTGTC. 1869
639 roAsrAspAspAlaGlnLysCysGlnleuIecysGlnSerAlaAspThr 655
1870 .....AATGAGTGAACCCCTCGAGTGCAGTCGCCGCCGCCGGAATGAG 1914
656 GluAspValAlaPheMetAsnGlnValValHisAspGlythrArgCys. 671
1915 TACTTTGCCGAGAGCTGCGGAGCGCGTGTGATGAGCAACCCCTGCTA 1964
672 ..SetIlyArgAspProTyrSerValCysAlaArgGlyGluCysValPro 688
1965 ccaggtccgacccagccgagcactctgcatcAACGCGCATCTGTAAAGAC 2014
688 alGlyCysAspLysGlnValGlySerMetLysAlaAspAspLysCysGly 704
2015 tggcgtgacttcagatgactcctgctgctatgagagacccgtgtgt 2064
705 ValCysGlyIlyAspAsnSerHisCysArgThrValLysGlyThrleuG 721
2065 gtgtgccacggcagcagcgtccacacccgacacccgagcagcagcagc 2114
721 ylysalaserlysglnalaglyAlaLeuLysLeuValGlnIleProAlaG 738

```

```

2115 GAGAGCC...GAGGGCCTGGGTATGTGATGTGGGCTGATCCAGCCG 2161
    ::||| ::||| ::||| ::|||
738 IyAlaArgHisIleGlnIleGluAlaLeuGlnLysSerProHisArgSer 754
    ||||| ::||| ::||| ::|||
2162 gccgacgcgagatccgcatccaaagaggttgcgagagcttgcacattctc 2211
755 ValValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGly 771
    ::||| ::||| ::||| ::|||
2212 GCACTGGGAGTGAAGACCCGGAGAACTACTTCTCATATGGTGGTGAC 2261
771 sglualatThrSerArgThrPheThrAlaMetGlyLeuGlnTrrGluAsp 788
2262 CATCCAGTGAACGGGACTACACAGTGGCAGGAGCCACCTTCACATACG 2311
788 IValGluAspAlaLysGlnSerleuLysThrSerGlyProleuProGlu 804
2312 CACGACGGGCACTGGAGAACCTCACTCCCGGCTCCACCAAGAG 2361
805 AlaIleAlaIleLeuAlaLeuProProThrGlnLysGlyProArgSer 821
2362 cctgtctgatccagctgctg...ttccagagagacacactt.....gg 2402
821 rleuAlaTyrLysTyrValIleHisGluAspLeuProleuIleGly 838
2403 ggtgcactacgagtaacacatccacagagag.....gcaggtg 2440
838 eAsnAsnValleuLeuLysGlnMetAspThrTyrGlnTrrAlaLeuLys 854
2441 gccacgacagagtc.....ccgcccgcctgtctcctgcacattatggg 2484
855 SerTrrAlaProCysSerLysAlaCysGlylyleuIleGlnPheThrly 871
2485 ccctggaccagatgcacagtcacctgcggcagagctgctgcagacagaa 2534
871 sTyrGlyCysArgArgArgAspHisMetValGlnArgHisleuc 888
2535 tctgtacgtcttgagcggcagcagcgcctgcagcagagacac...T 2581
888 ysAsnHisLysAspArgProLysProIleArgArgArgCysAsnGlnHis 904
2582 GTCACCCCTGGGCGCGCTGATGACCAAGAGAACTGCACGAGAG 2631
905 ProCysSerGlnProValTrrValThrGlnLysTrrGlyAlaCysSerAr 921
2632 ccttcctccctggcagc...TGTGGGCACTGAGTGCAGCTGCTCAG 2678
921 gSerCysGlyLysleuGlyValGlnThrArgGlyIleGlnCysleuLeuP 938
2679 ctcttcggcgctggggccttcctccggcgccgtgctgcattccgca 2728
938 roLeuSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaGly 954
2729 gcgtggcctgcatgacgacagccgcttcgacacccgctgttaacac 2778
955 ...AspArgProGluAlaArgArgProCysleuArg...ValProCysAr 969
2779 ctccccccgccccctactaAAACCCCTTCCACACCCCAATGACCTGTC 2828
969 oAlaGlnTrrArgLeuGlyAlaTrrSerGlnCysSerAlaThrCysGlyG 986
2829 ggccacactgggctgtgggaactgcttcagtcagtcagtcagtcagtc 2878
986 IuglyIleGlnGlnArgGlnValValCysArgTrrAsnAlaAsnSerleu 1002
2879 AGGCACTCAGCGCCGAATGTCCTTGC...ACCAATGACACCGGTGTC 2925
1003 GlnHisCysGlnLysArgProArgProAspThrValGlnLysSerleuP 1019
2926 ccctgtgacagagccacagcagcagcagcagcagcagcagcagcagc 2975
1019 oAlaCysGlyGlyAsnHisGlnAsnSerThrValArgAlaAspValTrrG 1036
    |||

```

2976 ACTCTGTCGC.....TGCC 2969
 1036 luleuGlyThr.....ProGluGlyGlnThrProGlnSerGly... 1049
 |||||
 2990 CCTTGGCGACACTGAGGCGCTGAAGGCTCAGGAGCGGCTCTCTCCAGCCAC 3039
 1050ProLeuHisProLeuAsnLysII 1057
 3040 GAGCTCTTCAACGAGCTGACTTATTCGCCGACACCTGCGCCACGCCCC 3089
 1057 eSerSerMetCysAlaIaGluProCysThrGlyAspSerValPheC 1074
 |||
 3090 TTCACCCGCTCATCACCCACGACGAGCACCATTGGCGAACGCAATTGAGG 3139
 1074 ysgInMetGluValLeuAspArgTyrCysSerIleProGly..... 1087
 :|||
 3140 AGAAGGCTCCAGAGCTGGAC.....CTCCGGGGGCCCGTGT 3177
 1088TyrHisArgLeuCysCysValSerCysIleLysIysAl 1100
 |||||
 3178 GTGAGCGACTTCTACTAGCACTACAATTTCATATTTCACAGAGATCT 3227
 1100 aSer...GlyProAsnProGlyProAsp..... 1108
 |||||
 3228 GTCTCTAGGGGCCCTCTGAGGAGCCGAGACTAGACCTGGCGGGACAGGG 3277
 1109ProGlyProThrSerLeuProProPheSerThrProGlySer 1122
 |||||
 3278 ACCGAGACCCCGACACACACCACTCTGTGGCGCCCTCAGCGGTAGC 3327
 1123 ProLeuProGlyLysProGlnAspProAlaAspAlaIaGlu.....Pr 1136
 |||||
 3328 CCGTGGCTGCCACAGAGGCTCTCTCCACCCAGAGGAGGAGGGGTACTGGG 3377
 1136 oProGlyLysProThrGlySerGluAspHisGlnHisGlyArg..... 1150
 |||||
 3378 ACCTTGTGTCGGGAGC...CGTGGGCTAGCGACGGCGCGCGTCCCCAC 3424
 1151AlaThrGlnLeuProGly...AlaLeuAspThrSerSerPro 1163
 :|||
 3425 CCCCACCTCTCAGAGCAGACCCCTGGGAGACCTTGATCATTTCTCGCT 3474
 1164 GLyThrGlnHisProPhe...AlaProGlnThrProIleProGlyLase 1179
 :|||
 3475 GAGGAGAGACACCCCATAGGGGCCCGACATCTTGGGCTCCCGAGCTGTC 3524
 1179 rTTPSer.....IleSerProThrThrPro.... 1187
 |||||
 3525 CTGGGCCACAGGTTTCCACTGATGGCTGACAGACACCTGCCACCCCTGAGA 3574
 1188GlyLeuLeuProThr 1192
 3575 GCCAAATGATTTCCAGTTGGCAGAGCAGCCAGCAGCAGCTGCCCT 3624
 1193 GLyTTPThrGlnThrProThrProValProGlnAspGlyGlnProGln 1209
 |||||
 3625 CCATGGCGGAGACAGACCAATGAGGTTTCAAGAGATGAGAGAACCC... 3672
 1209 yGluAspLeuArgHisProGlyThrSerLeuProAlaAspLeuProGlyA 1226
 3673AAGGCGCGGAGACACCC...CACGTGCCCGCA 3703
 1226 rGProProGlu.....ProCysHisProThrGlyThr 1236
 |||||
 3704 GACCCAGCTCCACGCTGCCCTTTGCTGCTGTGGAGC 3744
 seq_name: gb_pat:AX128433
 seq_documentation block:
 LOCUS AX128433 5353 bp DNA linear PAT 15-MAY-2001
 DEFINITION Sequence 2 from Patent WO0131034.
 ACCESSION AX128433
 VERSION AX128433.1 GI:14134924

[illegible]

```

169 TCGTGGACACCGGTTCCGAGTGCACGCGGGGGGGCTCTTCTGTTCCATC... 215
57 ValSerGlyProAlaAlaAlaSerAlaGlySerMetValValAspThrPr 73
215 ..... 215
73 oProThrLeuProArgHisSerSerHisLeuArgValAlaArgSerProL 90
215 ..... 215
90 euHisProGlyGlyThrLeuThrProGlyArgValGlyArgHisSerLeu 106
216 ..... GAGCTGTGGCCCGCCGACCTCGCAAGCGGATGTA 251
107 ..... TyrPheAsnValThrValPheGlyLy 115
252 TCTGTGGCCCGGAGACGCGCGCCGCTTCTACAGACTCAATACCGGGGGG 301
115 sGluLeuHisLeuArgLeuArgProAsnArgArgLeuValValProGlyS 132
302 CGAGCTGGCTTCAACCTGACCGCCCAACACAGACCTGCTGGCGCCGGCT 351
132 eraseValGluThrPngLysAspPheArgGluLeuPheArgGlnProLeu 148
352 TTGTGAGGAGACGCGGGCGCC...GGCGGCTTGGCGCCCGCCGCAATC 398
149 ArgGlnGlu.....CysValTyrThrGlyValThrGly.... 160
399 CGGGCCACACCGCGGCTGCACCTGTGGCGAGGTGCAGACCTGTA 448
161 MetProGlyAlaAlaValAlaHisSerAsnGlyAspGlyLeuAlaGlyL 177
449 GCTCAGAGGTGGCTCGGGCGGCATTCAGCCCTGCGACGCGCTGAAAGTG 498
177 euLeuArgThrAspSerThrAspPhePheLeuProLeuGluArgGly 193
499 TGTTCACACTCTCAACGAGGACTACTTCATTGAGCCCTGGAC...AGT 545
194 GlnGlnGluLysGluAlaSerGlyArgThrHisValThrArgArgG 210
546 GCCCGCGCGGCTGGCCACGCGCCACGCCCATGTGTGTACAAAGGTCA 595
210 uAlaValGlnGlnGluTTPAlaGluProAspGlyAspLeuHisAsnGluA 227
596 GCCCGCGGAGAGGCTGGCACAGCGGGGTATTTCAGTGTCCAAAGACT 645
227 lArPheGlyLeuGlyAspLeuProAsnLeuLeuGlyLeuValGlyAspGln 243
646 GT...GGAAGTGCAGGTACCAAGACTT... 671
244 LeuGlyAspThrGluArgArg... 251
672 .....GAGTCTGCAGCGGAGCGCTTGGAGACGCGGACGAGCGCGG 715
252 .....ArgHisAlaLysProGlySerTyrSer.....I 261
716 GCGACGCGGTGAGGCGCTGTACACCGAGCGGTGCGCAAGAAAGAAAGTGG 755
261 LeuValLeuLeuValAlaAspAspSerValAlaArgPheHisGlyLys 277
766 TGGAGACCGCTGATGAGTGCATGATGCCAAATATGGTGGAGTACCGAGAC 815
278 GluHisValAlaGlnAspArgValLeuThrLeuMetAsnLLeuAlaSprGlu 294
816 CGCGAGGTGAGAGCATATGGCTGACCATATGAAATCGTGGCT 865
294 eTyrHisAspGluSerLeuGlyValHisLeuAsnLLeuValValArgL 311
866 GTTTCATGACCGCAGCATGGGGAACCCACATCCATCCACTTGTGGCC 915
311 euLeuMetVal...GlyTyrArgGluSerLeuSerLeuLeuGluArgGly 326
916 TGTGCTGCTGGTGAAGATGAGAGAGAGGACTTAAGATCAGCGCACATGCA 965

```

327	AsnProSerAArgSerLeuGlnGlnValCysArgTrpAlaHisSerIleGln	343
966	GAC.....AACCCCTGGAAAGAGTTCTGGCAAGTGGACGAAAGCATCA	1009
343	narGlnAspProSerHisAlaGlnIleHisHisAspHisValValPheLeu	360
1010	CATGAAGGGGATGCGCATCCCTCCATGCACATGACAGTGCATCTGCTG	1059
360	hrrArgGlnAspPheGlyProSer.....Gly	368
1060	CGAGAAAGAGCTGTGTGGCAGCATGAAACCGGGCCCTGTGACACCTGGGA	1109
369	TrpAlaProArgLThrGlnMetCysHisAspProLeuArgSerCysAlaLeuAs	385
1110	CTGTGCCATGTGGGGGGATGTGGCAGCCGAGCCAGCTGCAGCATCAAA	1159
385	hHisGlnAspGlyPheSerSerAlaPheValIleAlaHisGlnThrGlyN	402
1160	CGAGAGACAGGGGCGTGGCGCTGCATCTGCATGAGCCACGAGCTGGGG	1209
402	IsaValLeuGlyMetGlnHisAspGlyGlnGlyIleAsnGlyCysAlaAspGlu	418
1210	ACAGTTTGGCATTTAGCATACGGAAGCGGCATGACTGATGAGCCGCT	1259
419	ThrSerLeuGlySerValMetAlaProLeuValIleAlaAlaPheHisAr	435
1260	GGGAAACAGCCTTTCATCATCTTCACAGCTCCTGTAGCAGCGCGCTCC	1309
435	gPheHisTrpSerArgCysSerLeuGlnLeuSerArgArgTrpPro.	451
1310	CCTCACCTGTGCTCCGCTGACCGCGCAGTATATACACAGATGCTTGACC	1359
452	..SerTrpAspCysLeuLeuAspAspProPheAspProAla.....	464
1360	GGGGTGGGGCTGTGGCTGGACGACCTT.....CTGGCAAGACAT	1403
465TrpProGlnProProGlnLeuProGlyIleAsnTrpSerMetAs	479
1404	ATGCAGCTTCCCTCGGTGCA.....CTGGCGCTCTGTATGATGTAAG	1447
479	psrGlnGlnAspArgPheAspPheGlySerGlyTrpIleThrCysLeuAlaP	496
1448	CCACGACGTCCGCTCCACAGTACGAG.....GCT	1476
496	hearGthrPhe.....GluProCysLysGlnLeuTrpCys	507
1477	AcCTGCTCCCTTGTGGCAGACATGATATATGTGTCACACACTGTGGTGC	1526
508	SerHisProAspAsnProTrpArgPheCysLysThrLysGlyProProLe	524
1527	TCGTGTGGGACCAAC.....TGTACTGCCAAGCTGATGACAGCCGT	1567
524	uAspArgLThrGlnCysAlaLarProGlyLysTrpCysPheLysGlnHisCys	540
1568	GGACGCGACCCGCTGTGGGGAATTAAGTGTGTCTCACTGGGAGATGCG	1617
541IleTrpLysSerProGlnGlnIleThrGlyGlnAspGlyGlyTrp	555
1618	TACCCGTGGGCTTCCGGCCGAGAGCCGTG.....GATGTGGCTGG	1658
556	SerSerTrpThrLysPheGlySerCysSerAlaArgSerCysGlyGlyGlyA	572
1659	TCGTGGTGGAGCCCTGTTCATCTGTGTCACGAGCATGTGGCATGGGCT	1708
572	LarGSerAlaArgSerArgSerCysAsnAsnProSerProhAlaTrpGlyGlyA	589
1709	ACAGACGCGCAGACCGGCACTGACGAGGACTTACGCCAAATPACAAAGCA	1758
589	rrpProCysLeuGlyProMetPheGlnTrpGlnValCysAsnSerGlnGlu	605
1759	GATACTGTGTGGGTAGGCCCAAGCGCTTCCGCTCTGTGCAACTGTGAAGCC	1808

606 CysProGlyThrTyGluAspPheArgAlaGlnIncLysAlaLysArgAs 622
 1809 TGCCCTGCTGGCCRCGCCCTCTTCCGCCACGTCCAGTGCAGCCACTTTGA 1858
 622 nSerTyTrpValHisGlnAsnAlaLysHisSerTrpValProTyrGluP 639
 1859 CGCT...ATGCTTAAAGAGGCCAGTGCACACATGGGTCCCGGTGTC. 1904
 639 roAspAspAlaGlnLysCysGlnLeuLeuLeuLeuSerAlaAspThr 655
 1905ATGAGGTGAACCCCTGGAGCTGCATGCCGCCGCCGGAATGAG 1949
 656 GlyAspValValPheMetAsnGlnValHisAspGlyThrArgGys... 671
 1950 TACTTGGCCGAGAGCTGGCGGAGCGCGTGTGATGGCAGCCCTGTCTGA 1999
 672 .SerTyArgAspProTyrSerValCysAlaArgGlyLysValPro 688
 2000 CCAGGTCCGAGCCAGCCCGGAGCTTCATCAACGCGCATCTGTAAAGACG 2049
 688 alGlyCysAspLysGlnValGlySerMetLysAlaAspLysCysGly 704
 2050 TGGGCTGTGACTTCGAGATTGACTCCGCTGATGAGAGAGCGCTGTGCT 2099
 705 ValCysGlyLysAspAsnSerHisCysArgThrValLysGlyThrLeuG 721
 2100 GTGGCCAGCGCAACGGCTCCACCTGCCACACCGTGAGCGGAGCTTCA 2149
 721 yLysAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnIleProAlaG 738
 2150 GGAGGCC...GAGGGCTGGGTATGTGTGATGGGGGTGATCCAGCCG 2196
 738 yLysAlaArgHisIleGlnIleGlnAlaLeuGlnLysSerProHisArgSer 754
 2197 GCGGACCGGAGATCCGATCCAAAGAGTTCGCCAGGCTCCCAACTTCCTG 2246
 755 ValValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGly 771
 2247 GCACGTGGAGTGAAGACCCGAGAGTACTTCCTCAATGCTGGGTGAC 2296
 771 sGluAlaThrSerArgThrPheThrAlaMetGlyLeuGlnIleProLys 788
 2297 CATCCAGTGAGAGGGGACTACCGAGTGGAGGACCACTTCACATAG 2346
 788 AlaValAlaAspAlaLysGlnSerLeuLysThrSerGlyProLeuProGlu 804
 2347 CACGACGAGGCAACTGGAGAACCTACGTCGCCGGGTCCCAACAGAG 2396
 805 AlaIleAlaIleLeuAlaLeuProThrGlnGlyProAlaGlySer 821
 2397 CCGTGTGATCCAGCTGCTG...TTCCAGAGAGCAACCTT...GG 2437
 821 rLeuAlaTyLysTyTrpValIleHisGlnAspLeuLeuProLeuIleGly 838
 2438 GGTGCACTACGAGTACACCATCCACAGGAG...GAGGTG 2475
 838 eTrAsnAsnValLeuLeuGlnIleMetAspThrTyGlnIleProAlaLeu 854
 2476 GCCACAGCAGAGTC...CCGCCGCCGTGTCTCTGCAATTAATGAG 2519
 855 SerTrpAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLys 871
 2520 CCCTGGACCAAGTGCACAGTCACTGCGGAGAGTGTGACAGAGAGAA 2569
 871 sTyTrpGlyCysArgArgArgArgPheHisMetValGlnArgHisLeu 888
 2570 TGTGACTGCTTGGAGCGGAGGAGGCGCTGACAGAGAGAC...T 2616
 888 yAspHisLysLysArgProLysProIleArgArgArgCysAsnGlnHis 904
 2617 GTGACCCCTGGGCGGCTGTGACCAACAGAGAAAGTGCAGGAGACG 2666
 905 ProCysSerGlnProValTrpValThrGlnIleTrpGlyAlaCysSer 921

2667 CCCGTCCCTGACAG...TGTGGGAGGTAGTGGAGCTGTGCTCCAG 2713
 921 gSerCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuLeuP 938
 2714 CTCCTCGGGGCGCTGGGAGCTTCGCCCGGCGCGCTGTGATCGGCA 2763
 938 roLeuSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaGly 954
 2764 GCGTGGGCTGGATGACAGAGGCGCTTGAGAGCACCCTGTGAAAC 2813
 955 ...AspArgProGlnAlaArgArgProCysLeuArg...ValProCysPr 969
 2814 CTCGCCGCGCCCTTACMAACCCCTTGCAACCGCGCATGTACCTGTC 2863
 969 oAlaGlnTrpArgLeuGlnValTrpSerGlnCysSerAlaThrCysGly 986
 2864 GGCACCTGGGCTGTGGGAGTGTCTGACTGCTGAGTGAATGTGGG 2913
 986 lGlyIleGlnGlnArgGlnValValCysArgThrAsnAlaAsnSerLeu 1002
 2914 AGGCACTCAGCGCGCAATGTCTCTG...ACCAATGACACCGGTTC 2960
 1003 GlyHisCysGlnLysAspArgProAspThrValGlnValCysSerLeuP 1019
 2961 CCTGTGACGAGGCCAGCAGCCAGCCAGCAAGTCACTCTCTGTCG 3010
 1019 oAlaCysGlyLysAsnHisGlnAsnSerThrValArgAlaAspValTrp 1036
 3011 ACTGTGCGG...TGGC 3024
 1036 lLeuGlyThr...ProGlnGlyIleTrpValProGlnSerGly... 1049
 3025 CCTGGGCACTGAGCGGCTGAAAGGTGAGGAGCGGCTCTCCAGCCAC 3074
 1050ProLeuHisProIleAsnLysI 1057
 3075 GACCTCTTACAGAGGCTACTTCATCCGACACCTGGGCCCCAGCC 3124
 1057 eSerSerMetCysAlaAlaGluProCysThrGlyAspArgSerValPheC 1074
 3125 TTCACCGGCTCATCCACAGCAGGAGCCATGGGCAACGCAATTGAGG 3174
 1074 ySlnMetGlnValLeuAspArgTyTrpCysSerIleProGly... 1087
 3175 AGGAGCTCCAGAGCTGAC...CTGCCGGGCGCGTGT 3212
 1088TyrHisArgLeuCysCysValSerCysIleLysLysAl 1100
 3213 GTGGAGACTTCTACTACAGTACATTTCAATTTTCCACGAGATCT 3262
 1100 aSer...GlyProAsnProGlyProAsp... 1108
 3263 GTCCTAGCGGCTCTGAGAGCCCGATGTAGACTGGCGGAGACAGGG 3312
 1109ProGlyProThrSerLeuProProPheSerThrProGlySer 1122
 3313 ACCGACACCCCCACACACACACCTGCTGCGCCCTTCACAGGGTAGC 3362
 1123 ProLeuProGlyProGlnAspProAlaAspAlaIleGlu...Pr 1136
 3363 CCTGTGCTGCCACAGAGCTCTGACACCAAGAGAGAGGGGACTGGG 3412
 1136 oProGlyLysProThrGlySerGlnAspHisGlnIleGlyArg... 1150
 3413 ACCTGTGTCGCCGAGC...CCTGGCTAGCCAGCGCGCGCTCCAC 3459
 1151AlaThrGlnLeuProGly...AlaLeuAspThrSerSerPro 1163
 3460 CCCGACCTCAAGCAGACCCCTGGGAAACCTTGTATCAATTTCTGCT 3509
 1164 GlyThrGlnHisProPhe...AlaProGlnThrProIleProGlyAla 1179
 ... 1188

```
3510 GAGGAAGACACCCCATAGGGGCCCCAGATCTGGGCTCCCCAGCCTGTC 3559
1179 fThpSer.....IleSerProThrThrPro.... 1187
      |||||
3560 CTGGCCCAAGGGTTTCCACTGATGGCTGCAGACACTGCCCCCTGAGA 3609
1188 .....GlyGlyLeuProTrp 1192
      |||||
3610 GCCAAATGATTTCCAGTTGGCAAGACAGCCAGCCAGCTGCCCCCT 3659
1193 GlyTrpThrGlnThrProThrProValProGlnsPlyGlyGlnProG1 1209
      |||
3660 CCATGGCGGGACAGACCAATGAGGTTCAGAGATGATGAGGAACCC.. 3707
1209 yGlnAspLeuArgHisProGlyThrSerLeuProAlaAspLeuProGlyA 1226
      |||
3708 .....AAGGGCGGGAGAACCC...CACCTGCCCCCGCA 3738
1226 rGProProGlu.....ProCysHisProThrGlyThr 1236
      |||||
3739 GACCACAGTCCACGCTGCCCTTGTGTCCTGTTGGCAGC 3779
```

	seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAS97178	
seq_documentation_block:		
ID AAS97178 standard; cDNA; 3561 BP.		
XX	AAS97178;	
AC		
DT 26-FEB-2002 (first entry)		
XX		
DE Human metalloprotease partial DNA sequence #7.		
KW	Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;	
KW	vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;	
KW	hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;	
KW	anorectic; antiinflammatory; aspartyl protease; cysteine protease;	
KW	metalloprotease; serine protease; cancer; haematopoietic; breast; colon;	
KW	lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;	
KW	immune-related disease; cardiovascular disease; neuronal disease;	
KW	migraine; sexual dysfunction; mood disorder; attention disorder;	
KW	cognition disorder; hypotension; hypertension; psychotic disorder;	
KW	dyskinesia; metabolic disorder; inflammatory disorder; ss.	
OS Homo sapiens.		
PN WO200183782-A2.		
PD 08-NOV-2001.		
PE 04-MAY-2001; 2001WO-US14431.		
PR 04-MAY-2000; 2000US-201879P.		
PS (SUCE-) SUGEN INC.		
PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;		
PT Payne V;		
DR WPI: 2002-041502/05.		
XX P-PSDB; AAU72895.		
XX Novel protease polypeptide useful for screening for substances that may		
PT be used to treat, e.g., cancers, immune-related diseases,		
PT cardiovascular disease, migraine, pain, psychotic and inflammatory		
PT disorders -		
PS Claim 30; Figure IK-L; 232pp; English.		
XX The invention relates to an isolated, enriched, or purified protease		
CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to		
CC screen for substances (S) that may modulate its activity. Administering		
CC S (which modulates protease activity in vitro) may be used to treat a		
CC disease or disorder selected from cancers (e.g., of tissues, of blood or		
CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,		
CC brain, ovarian, bladder or kidney), immune-related diseases and		
CC disorders, cardiovascular disease, brain or neuronal-associated diseases		
CC (e.g., central or peripheral nervous system diseases, migraine, pain,		
CC sexual dysfunction, mood disorders, attention disorders, cognition		
CC disorders, hypotension, hypertension, psychotic disorders, neurological		
CC disorders and dyskinesias), metabolic disorders and inflammatory		
CC disorders. (I) may also be useful as a diagnostic tool for a disease or		
CC disorder such as those above. AAs97159-AA97195 represent human		
CC protease coding sequences and primers of the invention.		
XX Sequence 3561 BP; 694 A; 1131 C; 1113 G; 623 T; 0 other;		
SQ alignment scores:		
/SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ32000 + 1098.50 750.71 1.8e-33		
/SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC90057 + 1098.50 750.71 1.8e-33		
/SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAK04374 + 1088.50 741.33 6.1e-33		
/SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ36862 + 1079.00 732.45 1.9e-32		
/SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF63440 + 1072.50 733.12 1.7e-32		

267	laspasbervvalvalarphenhsglyslglnhlsvalghnastlyr	284
2801	ggagagactcggtggctgcttccatggcgaagagacatgtccaaactatg	850
284	alleuthrlemetasnillevalasplululeyrhnsaspclusleu	300
851	ttcttacccttcataatgatgtacatgacgattttaccacgatgagtcctg	900
301	gllyalhisllleasnillealaleuvalargleuilemetvaldilyrhr	317
901	ggggcttcattatmaattatggcttgcctgtgcctgattcattggatgctacg	950
317	gglserseuenserleuilegiuarqglyasproserarqserleu	334
951	acatgctcctgagcctgactgacagccgggaaccccttcacgcagcctgagc	1000
334	lnvalcysargtrpalaahisserclelnarglnaspproserhlsala	350
1001	aggtgtgtcgctggcgacacacccccagacgcgccagaccacacgact	1050
351	gluhshlsasbphlsvalvalphleuthrarglnasphneglyprose	367
1051	gagcaccatgacacagtggtgttcttaccgccgagagactttggccctc	1100
367	rglytyralaprovaltrhrglymetcyshtsproleuargsercysala	384
1101	agggatgacacccgcactggcattgtgcacccccctgagagcgtgctgcc	1150
384	eulasnhsiglnaspllyhseaseralaephvalillealhsiglnthr	400
1151	ttcaccatgagagatggcttcttccacggcttgcgataagctcatgacac	1200
401	glyhlsvalleuglymetgluhlsaspllyglnlyasnelycysalaas	417
1201	ggccacagctcctggcatgagacatgacgctcaggggaatgctgtgcaga	1250
417	pcluthrserleuglyservalmetalaproleuvalglnalalaphen	434
1251	ttagaccacacccctggagacgctatgacccctctgctgacgctccttcc	1300
434	lsarbphehistrpsearqcyserlyslleugluuenserarqlyleu	450
1301	accgcttcattggtgccctgcacacagctgagacgctcagcgcctacctc	1350
451	prosertryaspdcysleuenuaspprobphasproalatrprogl	467
1351	ccctcctacgactgctcctctcatgacaccccttgcctgctgacccca	1400
467	nprrprogluleauproglylleasntryserketasplunglncysarq	484
1401	ggccccagagctgcttggsgatgaactactcaatgatatgacagatggcgt	1450
484	heasrphneglyserglytryglnthrlysleualaiphearqthpneglu	500
1451	ttgacttttggcagtgctaccacgacgctgctggcattcagagacctttgag	1500
501	procylysglnleuthrprocyserhspobaspanprotyrphcysly	517
1501	ccctgacaaacagctgtgggacagccattctgacaaacccgtacttctgcaa	1550
517	sthrlyslsglyproproleuaspolythrnglncysalaproglylvt	534
1551	gaccagaagaggggcccccgctgagatgggacactgagtgctacacccggcagt	1600
534	trpcysphelysglyhiscyslietrrplyssetproglnlththryrly	550
1601	gggtgcttcaaaagctctcatctcggaaatcccccagacagcatatggc	1650
551	glasnpgllyltrypsersertprltrylphedglysercyssearqse	567
1651	cagagatgagagctggagactcctcgaaacaaatttgggctatgtttcgcgcc	1700
567	rcysglyglylyvalargserarqserarqsercysasnsp-proserp	584

```

|||||
1701 ATGGGGGGGGGGTGCATCCCGCAGCGGAGCTGCACACACCCCTCC 1750
584 roAlaIyrgLygIyArpProCySLeuGlyPromePhegluTrgInVal 600
1751 CAGCCATATGAGGCCCGCGCTTAGGGCCATGTTCAGTACCAAGCTC 1800
601 CyASnSerGluGluGlyProGlyThrTrgIuAspPheAlaGlnG 617
1801 TGGAAACAGGAGAGTGCCTCGGACCTACGAGACTTCGGGCCACCA 1850
617 nCySAlaIyAspAsnSerTrpValHisGlnAsnAlaIySHisSer 634
1851 GTGTGCCAAGCCAACTCTACTATGTGACACAGATGCCAAGCAGACT 1900
634 rPValProTrgIuProAspAspAlaGlnIySLeuGluLeuIleCyS 650
1901 GGGTGCCTACAGCCTGACGATGACGCCCAAGACTGTGAGCTGATCTG 1950
651 GlnSerAlaAspThrGlyAspValAlaPheMetAsnGlnValHisAs 667
1951 CAGTGGCGGACACGGGGAGCTGTTCATGAAACAGGTGTTCACGA 2000
667 pGlyThrArpCySserTrpArpAspProTrpSerValCySAlaArpGly 684
2001 TGGGACAGCTGCAGCTACCGGACCCATACAGCTGTGCGCGTGGCG 2050
684 lCySValProValGlyCySAspIySgIuValGlySerMetLysAlaAsp 700
2051 AGTGTGCTCTGCTGCTGTGACAAAGAGGTGGGTCCATGAAGCGCAT 2100
701 AspIySgIyValCySgIyGlyAspAsnSerHisCySArpThrValIy 717
2101 GACAAGTGTGGAGTCTGCGGGGTGACAACTCCACAGCAGAGACTGGA 2150
717 sGlyThrLeuGlyLysAlaSerIySgInAlaGlyAlaLeuLysLeuVal 734
2151 GGGGACGCTGGCAAGCCTCCAAAGCAGGCGAGCTCAAGCTGTGTC 2200
734 lInIleProAlaGlyAlaArpHisIleGlnIleGluAlaLeuGluLysSer 750
2201 AGATCCGACGAGGTGCCAGGCATCCAGATTGAGGCACCTGGAGAACTCC 2250
751 ProHisArpSerValIyAsnGlnValIleGlySerPheIleLeuAs 767
2251 CCCACGCGCATTTGTGTGAAGAACAGCTCACCGGACGTTCACTCCCAA 2300
767 nProLysGlySgIuAlaThrSerArpThrPheThrAlaMetGlyLeuG 784
2301 CCCCAAGGCGCAAGGAGCCCAAGCCGACCTTCACCGCATGGCGCTCG 2350
784 lUTrpGluAspAlaValGluAspAlaLysGluSerLeuLysThrSerGly 800
2351 AGTGGAGAGATGCGGTGAGAGATGCCAAGAAAGCCTCAAGACACAGGG 2400
801 ProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGluGly 817
2401 CCTGCTGCTGAAGCATTGCCATCTGCTCTCCCTCCCAACTGAGAGGTG 2450
817 yProArpSerSerLeuAlaTrpLysTrpValIleHisGlnAspLeuLeu 834
2451 CCCCCGAGAGCCTGGCTTACAAAGTACATCCATGAGAGACTGTCTGC 2500
834 rGluLeuIleGlySerAsnAsnValLeuLeuGluGluMetAspThrTrpGlu 850
2501 CCTTATCGGGAGCAACATATGCTCTCTGAGAGATGACACACATGAG 2550
851 TrpAlaLeuLysSerTrpAlaProCySserLysAlaCySgIyGlyIle 867
2551 TGGGCGCTCAAGAGCTGGGCCCTCTGACCAAGGCTGTGGAGAGGAT 2600
867 eGlnPheThrLysTrpGlyCySArpArpArpArpHisIleMetValG 884
|||||

2601 CCAGTTCCACAAATACGCTGCCGGCCAGAGAGACCACCAATGTCG 2650
884 lInArpHisLeuCySAspHisLysLysArpProLysProIleArpArp 900
2651 AGCGACACCTGTGTGACCAAGAAAGAGCCCAACCATCTCGCGGGCC 2700
901 CyASnGlnHisProCySserGlnProValTrpValThrGluTrpGlu 917
2701 TGGAAACAGCACCCTGCTCTCAGCCTGTGTGGTGAAGAGAGTGGGG 2750
917 yAlaCySserArpSerCySgIyLysLeuGlyValGlnThrArpGlyIle 934
2751 TGGCTGACACCGGAGACTGTGGGAAGCTGGGGGTGCAGACAGCGGGATAC 2800
934 lCySLeuLeuProLeuSerAsnGlyThrHisLysValMetProAlaLys 950
2801 AGTGCCTGATGCCCTCTCTCAATGGAACCCACAAAGTCACTCGGCCAA 2850
951 AlACySAlaGlyAspArpProGluAlaArpArpProCySLeuArpValPr 967
2851 GCCTGTGCGGGAGCCGCTGAGGCGCCGAGCGCCCTGTCTCCGAGTGC 2900
967 oCySProAlaGlnTrpArpGluGlyAlaTrpSerGlnCySserAlaThr 984
2901 CTGCCAGCCAGTGTGAGGCTGGAGCCTGGTCCAGTGTCTGCCACT 2950
984 ySgIyGluGlyIleGlnGlnArpGlnValAlaCySArpThrAspAlaAsn 1000
2951 GTGGAGAGGCGATCCAGCAGCGCAGGTGTGTGACAGACCAACGCCAAC 3000
1001 SerLeuGlyHisCySgIyLysArpArpProAspThrValGlnValCyS 1017
3001 AGCTCGGGCATGTGGAGGGGATAGCCAGACTGTCCAGGCTGTGACG 3050
1017 rLeuProAlaCySgIyGlyAsnHisGlnAsnSerThrValArpAlaAsp 1034
3051 COTGCTGCTGTGTA..... 3066
1034 aLTrpGluLeuGlyThrProGluGlyInTrpValProGlnSerGlyPro 1050
3066 ..... 3066
1051 LeuHisProIleAsnLysIleSerSerMetCysAlaAlaGluProCysH 1067
3067 ..... 3080
1067 rGlyAspArpSerValPheCySgInMetGluValLeuAspArpTrpCyS 1084
3081 GGGAGACAGGCTGTCTTGTGCCAGATGGAAGTCTGCATCCGCTACTGCT 3130
1084 erIleProGlyTrpHisArpLeuCySValSerCysIleLysLysAla 1100
1101 CCATTCGCGGTACACAGGCTGTGTGTGTCTGTGCATCAAGAAAGGCC 3180
1101 SerGlyProAsnProGlyProAspProGlyProThrSerLeuProProH 1117
3181 TCGGGCCCAACCTGGCCAGACCTGGCCCAACTCACTGCCCCCTT 3230
1117 eSerThrProGlySerProLeuProGlyProGlnAspProAlaAspAla 1134
3231 CTCACACTCTGGAAGCCCTTACCAAGAGCCCAAGACCTTGCAGATGCTG 3280
1134 lAgIuProProGlyLysProThrGlySerGluAspHisGlnIleGlyArp 1150
3281 CAAAGCCTCTGGAAGCCCAAGCGGATCAGAGACCATCAGCATGCGCGA 3330
1151 AlaThrGlnLeuProGlyAlaLeuAspThrSerSerProGlyThrGlnH 1167
3331 GCCACACAGCTCCCAAGAGCTGTGATACAACTCCCAAGGAGACCAACA 3380
1167 sProPheAlaProGluThrProIleProGlyAlaSerTrpSerIleSer 1184
3381 TCCCTTTGCCCTGAGACCAATCCCTGAGACATCTGAGACATCTCC 3430

```

1184 rothrhprproglvglyleuprotprgltyrthrghlthrprothrpro 1200
 3431 CTACACACCCCGGGGGCTGCTGGGGCTGACACACACACACGCA 3480
 1201 ValProGluAspLysGlyInProGlyGluAspLeuArgHisProGlyTh 1217
 3481 GTCCCTGAGAGCAAGGGCAACTGGAGAACTGACATCCGCGAC 3530
 1217 rSerLeuProAlaAspLeuPro 1224
 3531 CAGCTCTCCCTGCTGCTCCCGC 3552

seq_name: /STD5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH47048

seq_documentation_block:
 ID AAH47048 standard; DNA; 3144 BP.

XX AC AAH47048;

29-OCT-2001 (first entry)

Larynx carcinoma associated protein-1 (LarCAP-1) encoding DNA.

Larynx carcinoma associated protein-1; LarCAP-1; carcinoma; cytostatic;
 anti-metastatic; anti-arthritic; osteopathic; immunosuppressive; human;
 immunostimulant; cerebroprotective; vasotropic; anti-inflammatory;
 dermatological; cardiant; vaccine; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 1..3144

FT /tag= a
 /product= "LarCAP-1"
 /note= "the stop codon is not indicated"

XX FN WO200159133-A1.

XX PD 16-AUG-2001.

XX PF 12-FEB-2001; 2001WO-EP01525.

XX PR 14-FEB-2000; 2000EP-0102955.

XX PA (MERE) MERCK PATENT GMBH.

XX PI Duecker K, Hentsch B, Hoheisel J, Frohne M;

XX DR WPI: 2001-483569/52.

XX P-PSDB: AAB85695.

Larynx carcinoma associated protein-1 polypeptide for the treatment of
 carcinomas, metastasis, arthritis, osteoporosis, immune disorders,
 stroke, ischemia, autoimmune disease, angiogenesis, skin disorders and
 organ malfunctions -

XX Claim 5; Page 33-39; 47pp; English.

XX This DNA encodes a human Larynx carcinoma associated protein-1 (LarCAP-1)
 polypeptide. The LarCAP-1 polypeptide can be expressed by standard
 recombinant methodology. The LarCAP-1 polypeptide and polynucleotide are
 useful in diagnostic assays and for the treatment of carcinomas,
 metastasis, arthritis, osteoporosis, immune disorders, stroke, ischemia,
 autoimmune disease, angiogenesis, skin disorders and organ malfunctions
 especially heart hypertrophy.

XX Sequence 3144 BP; 633 A; 1000 C; 957 G; 554 T; 0 other;

alignment_scores:

Quality: 5160.50
 Ratio: 5.573

Length: 970
 Gaps: 2

Percent Similarity: 95.464 Percent Identity: 95.361

alignment_block:

..

Align seg 1/1 to: AAH47048 from: 1 to: 3144

291 ValAspGluIleuThrHisAspGluSerLeuGlyValHisIleAsnIleAl 307
 238 GTAGATGAGATTATCCACGATGATGCTGGGGGTTCTATTAATATGTC 287
 307 aleuValArgLeuIleMetValGlyTyrArgGlnSerLeuSerLeuIleG 324
 288 CCTGTCGCGCTTATCATGTTGGCTACCGACAGTCCCTGAGACCTGATCG 337
 324 LuArgGlyAsnProSerArgSerLeuGluGlnValCysArgTyrAlaHis 340
 338 AGCGGGGAAACCCCTCAGCAGCTGAGACAGGTGTGTGGGGGACAC 387
 341 SerGlnGlnArgGlnAspProSerHisAlaGluHisAspHisValVa 357
 388 TCCAGACAGCGCCAGAGACCCAGCCAGCTGAGACCATGACACAGCTTGT 437
 357 LpheLeuThrArgGlnAspPheGlyProSerGlyTyrAlaProValThrG 374
 438 GTTCTCAGCCCGGAGAGACTTTGGCCCTCAGGGATGACCCGCTACCTG 487
 374 LysMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAspGlyPhe 390
 488 GCATGTGTACACCCCTGAGAGACTGTGCTCCCTCAGCATGAGATGGCTTC 537
 391 SerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGlyMetG 407
 538 TCTTCAGCCTTTCGATGATGCTCATGAGACCGCCAGCTCTCGGCAATGA 587
 407 uHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuGlySerY 424
 588 GCATGACGGTCAAGGGAATGGCTGTGCAGATGAGACCACTGGGACGG 637
 424 aMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTyrSerArg 440
 638 TCATGGCCGCCCTGCTGTCAGGCTGCTCCACCGCTCATTTGTGCTCCG 687
 441 CysSerLysLeuGluLeuSerArgTyrLeuProSerTyrTAspCysLeu 457
 688 TGCAGCAAGCTGGAGCTCAGCCGCTACCTCCCTCAGACAGTGCCTCCT 737
 457 uAspAspProPheAspProAlaTyrProGlnProProGluLeuProGlyI 474
 738 CGATGACCCCTTTGATCTGCTGCTGGCCCAAGCCCAAGAGCTGCTGGGA 787
 474 LeuAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerGlyTyr 490
 788 TCAACTACTCAATGATGATGAGCAGTGCCTTGTGACTTTGGCAGTGGCTAC 837
 491 GlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeuTyrCy 507
 838 CAGACCTGCTTGGCACTTCAGACCTTTGAGACCTTCGCAAGCAGCAGTGG 887
 507 sSerHisProAspAsnProTyrPheCysLysThrLysLysGlyProProL 524
 888 CAGGCATCTCTGACACCCGCTACTTCTGCAGACCAACAAAGGGGGCCCCG 937
 524 euAspGlyThrGlnCysAlaProGlyLysTyrPheLysGlyHisCys 540
 938 TGGATGGAGCTGATGTGCACCCGCAAGTGTGTCTCAAAAGTCACTGC 987
 541 IleTyrLysSerProGluGlnThrTyrGlyGlnAspGlyLysTyrSerSe 557
 988 ATCTGGAATGTCGCGAGACAGACATATGCGCAGATGAGGCTGAGACTC 1037
 557 rTTPThrLysPheGlySerCysSerArgSerCysGlyGlyValAlaArg 574

```

1038 CTGAGCAAGTTTGGTCATGTTGCGCGTCATGTGGGGCGGGGTGCAT 1087
574 eRASeArGSeRSeCyAsnAsnProSeProAlaTyrGlyGlyAlaPro 590
1088 CCGGACCGCGAGGTGCAACACCCCTCCCGACGCTATGGAGCCCGCCG 1137
591 CysLeuGlyProMetPheGlyTyrGlnValCysAsnSerGlyGlyCysPr 607
1138 TGCTTAGGCCCCATGTTGAGTACCAAGCTCTGCAACAGCAGAGAGTCC 1187
607 oGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgAsnSer 624
1188 TGGACCTTACAGGACTTCGGGGCCAGAGTGTGCAAGCCCACTCTCT 1237
624 YrTYrValHisGlnAsnAlaLysHisSerTrpValProTyrGluProAsp 640
1238 ACATGTGCACCAAGATGCCAACACAGCTGGTGGCTTACGAGCCTGAC 1287
641 AsPAsPAlaGlnLysCysGlyLeuLeuLecysGlnSerAlaAspThrGlyAs 657
1288 GATGACGCCCGAAGTGTGAGCTGATGCTCCAGTCGGCGACACGGGGGA 1337
657 pValValPheMetAsnGlnValValHisAspGlyThrArgCysSerTyrA 674
1338 CGTGGCTTCATATACCAAGGTGTTGAGTGGAGACGCTGCACCTAC 1387
674 rGAsPProTyrSerValCysAlaArgGlyGlyCysValProValGlyCys 690
1388 GGGACCCATACAGCGTCTGTGCGCGTGGGAGTGTGCTGTCTGCGCT 1437
691 AsPlyGlnValGlySerMetLysAlaAspLysCysGlyValCysGly 707
1438 GACAAAGAGGTGGGTCCATGAAAGCGCATGCAAGTGTGAGTCTGCGG 1487
707 yGlyAsPAsnSerHisCysArgThrValLysGlyThrLeuGlyLysAla 724
1488 GGGTGACACTCCCACTGCAGAGACTGTAGAGGGAGCGCTGGGCAAGCT 1537
724 eTlyGlnAlaGlyAlaLeuLysLeuValGlnIleProAlaGlyAlaArg 740
1538 CCAAGAGGAGAGGAGCTCTCAAGCTGTGCGAGATCCAGCAGGTGCCAG 1587
741 HisIleGlnIleGlnAlaLeuGlnLysSerProHisArgSerValVal 757
1588 CACATCCAGATGAGGACCTGGAGAGATCCCGCACCGCATTTGGTGAA 1637
757 sAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGlnAla 774
1638 GAACCAAGTCACCGGAGCTTCATCTCAACCCCAAGGCGCAAGAACCA 1687
774 hSSeArGThrPheThrAlaMetGlyLeuGluTyrGluAspAlaValGlu 790
1688 CAGGCGGAGCTTCACCGCATGGGCGCTGGAGTGGAGATGGCGTGAG 1737
791 AsPAlaLysGlnSerLeuLysThrSerGlyProLeuProGlnAlaIleAl 807
1738 GATGCCAAGGAAGCTCTCAAGACCGGGCGCTGCTGTAAGCCATATGC 1787
807 aLleLeuAlaLeuProProThrGlnGlyGlyProArgSerSerLeuAla 824
1788 CATCTGCTCTCCCGCCCACTGAGGGTGGCGCCCGCAGCAGCAGCTGC 1837
824 YrLYSTYrValIleHisGluAspLeuProLeuLecysSerAsnAsn 840
1838 ACAAGTACGTCATCATGAGGACTGCTGCGCTTATCGGGAGCAACAA 1887
841 ValLeuLeuGlnGluMetAspThrTyrGluTyrAlaLeuLysSerTrpAl 857
1888 GTGCTCCGAGAGATGAGACCTATGATGGCGCTCAAGAGCTGGCG 1937
857 aProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTyrGlyC 874
1938 CCGCTGAGCAAGGCGTGTGAGAGGAGCATCAGTTACCAATAATAGCG 1987

```

```

874 ySArGArGArGArGAsPHisHisMetValGlnArgHisLeuCysAsPHis 890
1988 GCGGCGCAGACGAGACACACATGATGTGAGGAGACACTGTGTGACAC 2037
891 LysLysArgProLysProIleArgArgCysAsnGlnHisProCysSe 907
2038 AAGAAGAGGCCAAGCCATCCCGCGCGCTGCAACCAACACCCCTGTC 2087
907 rGlnProValTyrValThrGluTyrProGlyLysAlaCysSerArgSerG 924
2088 TCAGCCCTGTGGGTGACGAGAGAGTGGGTGCTGCGACCGCGAGCTGG 2137
924 LysLeuGlyValGlnThrArgGlyIleGlnCysLeuLeuProLeuSer 940
2138 GGAAGCTGGGGGTGCAGACACGGGGGATACAGTGCCTGCTGCCCTGCC 2187
941 AsnGlyThrHisLysValMetProAlaLysAlaCysAlaGlyAsPArg 957
2188 AATGGAACCCACAAAGTCAATGCCGCGCAAGCCTCGCGGGAGCCGCC 2237
957 oGlnAlaArgArgProCysLeuArgValProCysProAlaGlnTyrArg 974
2238 TGAGGCGCGAGCGGCTGTCTCGAGTGGCTGCCACAGCTGGAGGC 2287
974 eugLYAlaTrpSerGln..... 979
2288 TGGAGCCGTGGTCCAGAAATCTTGAGCAGCTAGCTGATGCCAGAC 2337
979 ..... 979
2338 CTGTGCTAAGATGAGGAGACCTAGCATCAACGACAGAGCTGATCT 2387
980 .....CysSerAlaThrCysGlyGlyGly 988
2388 TGCCCTGTCGACGCCACAGTCTGTGCTGTGCACTGTGGAGGGGCA 2437
988 IeGlnGlnArgGlnValValCysArgThrAsnAlaAsnSerLeuGlyHis 1004
2438 TCCAGCAGCGGCGAGGTGGGTGTCAGAGCAACGCGCAACAGCTGGGCAT 2487
1005 CysGlnGlyAsPArgProAspThrValGlnValCysSerLeuProAla 1021
2488 TCGAGAGGGGATAGGCGACAGCTGTCCAGTGTGCACTGCCCGCTG 2537
1021 sGlyGlyAsnHisGlnAsnSerThrValArgAlaAsPValTyrProL 1038
2538 TGGAGGAATCACCGAACTCCAGGTGAGGGCGCATGTGGAACCTTG 2587
1038 LYThrProGlnGlyGlnTyrValProGlnSerGlyProLeuHisProIle 1054
2588 GGAGCCCGAGAGGGAGTGGGTGCCACAAATCGAACCCCTACATCCCAT 2637
1055 AsnLysIleSerSerMetCysAlaAlaGluProCysThrGlyAsPArg 1071
2638 AACCAAGATATCATCA.....ACGAGCGCTGCAAGGAGAGACAGT 2678
1071 rValPheCysGlnMetGluValLeuAsPArgTyrCysSerIleProGly 1088
2679 TGTCTTGTGCAGATGAAGTGTGATCTGCTACTGCTCATTCGCCGGCT 2728
1088 YrHisArgLeuCysCysValSerCysIleLysLysAlaSerGlyProAsn 1104
2729 ACCACCGGCTGTGCTGTGTCTGCTGCAATCAAGAAAGCCTGGGCCCAAC 2778
1105 ProGlyProAsPProGlyProThrSerLeuProProPheSerThrPro 1121
2779 CTTGGCCAGACCTGGCGCAACCTGACCTGCCCTTCTCCACTGCTCTG 2828
1121 YSerProLeuProGlyProGlnAsPProAlaAsPAlaGluProProG 1138
2829 AAGCCCTTACAGAGACCCAGAGACCTGAGATGCTGAGAGGCTCTG 2878

```

```

1138  lylyspProthnglySerGluAspHisGlnHisGlyAlaThrGlnLeu 1154
1139  |||||||||||||||||||||||||||||||||||||||||||||||
2879  GAAGGCCAGCGGATCCAGAGGACCATGACATGCGCGAGCCAGACAGC 2928
1155  ProGlyAlaLeuAspThrSerSerProGlyThrGlnHisProPheAlaPr 1171
1156  |||||||||||||||||||||||||||||||||||||||||||||||
2929  CCAGGAGCTGTGATACAGCTCCCGGAGGAGCCAGCATCCCTTTGCCCC 2978
1171  cGluThrProThrProGlyAlaSerTrpSerTrpSerProThrThrProG 1188
1172  |||||||||||||||||||||||||||||||||||||||||||||||
2979  TGAGACACCAATCCCTGAGCATCTTGAGCATCTCCCTACACACCCCG 3028
1188  lyGlyLeuProTrpGlyThrThrGlnThrProThrProAlaProGlyAsp 1204
1189  |||||||||||||||||||||||||||||||||||||||||||||||
3029  GGGGGCTGCTTGGGGCTGGAGCTGAGACACCTGAGCCAGCTCCGAGGAC 3078
1205  lysGlyGlnProGlyGlyAspLeuArgHisProGlyThrSerLeuProAl 1221
1206  |||||||||||||||||||||||||||||||||||||||||||||||
3079  AAAGGCAACCTGGAGAGACCTGAGACATCCGCGACACGCTCCCTGCG 3128
1221  AspLeuPro 1224
1222  |||
3129  TCCCTCCCG 3138

```

seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAH20225

seq_documentation_block:

ID AAH20225 standard; cDNA; 3618 BP.

XX AC AAH20225;

XX D7 07-AUG-2001 (first entry)

XX DE Human ADAM-type metalloprotease MDT5-encoding cDNA, SEQ ID NO:11.

XX KW Human; MDT5; ADAM-type metalloprotease; drug screening;

XX KW A Disintegrin And Metalloprotease; cancer; arthritis; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..3618

XX FT /*tag= a

XX FT /product= "Human MDT5"

XX FT /note= "ADAM-type metalloprotease"

XX PN JP2001017183-A.

XX PD 23-JAN-2001.

XX PF 09-JUL-1999; 99JP-0196584.

XX PF 09-JUL-1999; 99JP-0196584.

XX DR WPI: 2001-275950/29.

XX DR P-PSDB: AAB73550.

XX PT A new metal protease and its preparation for use as an anti-cancer and

XX PT anti-arthritis therapeutic -

XX PS Example 2; Page 19-20; 22pp; Japanese.

XX PS The invention relates to the novel human ADAM (A Disintegrin And

XX CC Metalloprotease)-type metalloproteases MDT54 (AAB73549) and MDT55

XX CC (AAB73550). The metalloproteases can be used for the treatment of

XX CC cancers and arthritis. The invention also relates to the genes encoding

XX CC MDT54 and MDT55, vectors and host cells containing the MDT54 or MDT55

XX CC genes, the recombinant production of MDT54 and MDT55, and antibody

XX CC specific for MDT54 or MDT55, and methods of screening for compounds

XX CC which modulate the activity of MDT54 and/or MDT55. The present

XX CC sequence represents cDNA encoding human MDT55.

```

XX SQ Sequence 3618 BP, 1035 A; 802 C; 880 G; 901 T; 0 other;
alignment_scores:
    Quality: 3624.00      Length: 1198
    Ratio: 4.081          Gaps: 17
    Percent Similarity: 74.124      Percent Identity: 55.593
alignment_block:
us-09-938-330-20 x AAH20225
Align seg 1/1 to: AAH20225 from: 1 to: 3618

```

```

37 AspTyrGlyValThrValProCysSerThrAspPheArgGlyArgPhe 53
38  .... ValAlaArgSerProLeuHisProGlyGlyThrLeuTrpPr 98
208 GCGAGGAGCGTGTCTCCAACTT..... 231
98  oGlyArgValGlyArgHisSerLeuThrPheAsnValThrValPheGly 115
232  .... GAGCAGTTGTTCTTTAACTACATCCAGCGCATTTGGAA 265
115  yGlyLeuHisLeuArgLeuArgProAsnArgArgValValProGly 131
266  AAGATTTCATCTGCGACTTAAGCCCAACACTGACATGATGCTCTGGG 315
132  SerSerValGlyTrpGlnGlu..... 138
316  GCTGTGTGGAGTGGCATGACATCTGTGCTGCTGGAATTAACCGA 365
139  .... AspPheArgGlyLeuPheA 145
366  TCCCATTAACAACCAACCAAGAGTGCATGATGATGATGCGGAAA 415
145  rgGlnProLeuArgGlnGlyValTyrThrGlyValThrGlyMet 161
416  CAGAGCTTTGCGACACTAAGTGTGTTGTTGTTGTTGTTGTTGTTGTT 465
162  ProGlyAlaAlaValAlaIleSerAsnGlyAspGlyLeuIle 178
466  CCAGCAACCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 515
178  eArgThrAspSerThrAspPheIleGluProLeuGlnArgGlyGln 195
516  AAAAGTGTATATGAAGTATTTGATTTGATTTGATTTGATTTGATTT 565
195  InGlyGlyAlaIleSerGlyArgThrHisValValTyrArgGlyAla 211
566  AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 615
212  ValGlnGlnGlnTrpAlaGlnProAspGlyAspLeuHis.....As 225
616  GTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665
225  ngAlaIlePheGlyLeuGlyAspLeuProAsnLeuGlnGlyValGly 242
666  GAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 715
242  spGlnGlyAspThrGlyArgGlyArgGlyHisAlaGlyProGlySer 258
716  AGCAGTGAATGAACAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 765

```



```

259 TyrSerIleGluValLeuLeuValAlaAspSerValIleArgPheI 275
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
766 TACAATATGAGGTACTCTGGGAGTGAAGTACTGTGGTCCGTTTCCA 815
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
275 sglYysGluHisValGlnAsnTyrValLeuThrLeuMetAsnIleValA 292
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
816 TGGCAAGAAGCAAGTCACAAACTACCTCTGACCTTAATGAACATTGGA 865
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
292 spGluIleTyrHisAspGluSerLeuGluValHisIleAsnIleAlaLeu 308
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
866 ATGAATTTACATGATGAGTCCCTCGAGAGTCATATTAATGTGGTCTG 915
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
309 ValArgLeuIleMetValGlyTyrArgInsLeuSerLeuIleGluArg 325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
916 GTGGCATGATATGCTGGGATATGCACAAAGTCACACCTCATAGAAG 965
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
325 gglYAsnProSerArgSerLeuGluGlnValCysArgTyrAlaHisSerg 342
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
966 GGGAAACCATCCAGACCTTGAGAAATGTGTGCTGGCGCTGCCAAC 1015
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
342 IngIArgGlnAspProSerHisAlaGluHisAspHisValAlaPhe 358
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1016 AGCAAAAGATCTGATCTCAACACCTGTGACACACATGACCATGCAATT 1065
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
359 IeuthrArgGlnAspPheGlyProSer.....GlyTyrAlaProva 372
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1066 TTAACCAAGCAAGACTTTGGAACCTGTGATGCAAGATATGCTCCAGT 1115
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
372 IThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnIleGlnAspG 389
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1116 CACCGGCAATGTGTCTCCAGTGAAGAGTTGTACCTCAATCATGAGATG 1165
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
389 IyPheSerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 405
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1166 GTTTTTCATCTGCTTTTCTTACTACCCATGAACGGCCCATGTGTGGGA 1215
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
406 MetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuG 422
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1216 ATGGAGCATGTATGAGCAAGCAACAGGTGTGATGACAGACTGTATGGG 1265
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
422 ySerValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrps 439
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1266 AAGTGTCAATGGCTCCCTGTGGACAGACAGATCCATTCCTTGTGACTGT 1315
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
439 eArgCysSerLysLeuGluLeuSerArgTyrLeuProSerTyrAspCys 455
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1316 CCCGATGAGTGTCAAGAACTGAAAGATATATCCATTCTTGTGACTGT 1365
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
456 LeuLeuAspAspProPheAspProAlaTyrProGluProGluLeuArg 472
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1366 CTCTTATGACCCCTTTGATCATGTATGGCTTAAACTCCAGAACTTCC 1415
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
472 oGlyTlAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerg 489
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1416 TGGATATCAATTAATCTATGATGATGCAATGTCTTTGATTTTGGTGGT 1465
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
489 IyTyrGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeu 505
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1466 GCTATATAAAGTGCACCGCTTCCGAACCTTGACCCATGTAACAGCTG 1515
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
506 TrpCysSerHisProAspAsnProTyrPheCysLysThrLysLysGlyPr 522
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1516 TGGGTATGACCTCTGATTAATCCCTACTTTTGTAAAGACTAAAGAGGAC 1565
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
522 oProLeuAspGlyThrGluCysAlaLarProGlyLysTrpCysPheLysGlu 539
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1566 TCCACTTGATGGAGCTGAATGTCTCTGGAATAATGTCTTAAGAGTCT 1615
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
539 IScyIleTrrPlySerProGluGlnThrTyrGlyGlnAspGlyLysTrp 555
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1616 ATTGCAATGTGGAAGAAATGCTAATCAG...CAAAACAGATGACCAATTGG 1662

```

```

556 SerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyValA 572
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1663 GGGTCATGAGCACTAAATTGGCTCTCTGTCGGACATGTGGAACTGGTGT 1712
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
572 IArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGlyGlyA 589
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1713 TCGTTTCAGAACACCGCCAGTGCATTAATCCATGCTCAATGATGTGTC 1762
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
589 rGProCysLeuGlyProMetPheGluTyrGlnValCysAsnSergGlu 605
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1763 AGGATGTCTCTGTGCTTAATTTTGTAGTACAGCTTTGTAAACAGAAACA 1812
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
606 CysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgAs 622
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1813 TGGCAAAAACACTTTGAGAGCTTCAGACACAGCAGTGTGACGAGGAAA 1862
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
622 nSerTyrTyrValHisGlnAsnAlaLysHisSerTrpValProTyrGluP 639
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1863 CTCCCACTTTGATATCCAGAAATACCAACACCACTGCTGTGCTATGAT 1912
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
639 roAspAspAlaGlnLysCysGluLeuIleCysGlnSerAlaAspThr 655
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1913 ATCTGACACCCCAAGAAAGATGCCACCTTTACTGTCAGTCCAAGAGACT 1962
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
656 GlAspValValPheMetAsnGlnValValHisAspGlyThrArgCysSe 672
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1963 GGAGATGTTGCTTACATATAAACCACTGTGCTATGATGAACGACGTGTC 2012
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
672 IyTyrArgAspProTyrSerValCysAlaArgGlyGlnCysValProValG 689
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2013 TTACAATAATCCATATATACATATGTGTCCAGAGAGATGTGTGAATGTG 2062
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
689 IyCysAspLysGluValGlySerMetLysAlaAspLysCysGlyVal 705
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2063 GCTGTGATTAAGAAATGTGTTCTTAATAGGTGTGAGATAGTGTGGTTC 2112
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
706 CysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGly 722
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2113 TGTGGAGGAGATTAATTCCTGCTGCCAAGCTGAAGGGGACATTTACAG 2162
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
722 sAlaSerLysGlnAlaGlnAlaLeuLysLeuValGlnIleProAlaGlyA 739
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2163 AACTCCAGAGAAAGCTTGGTACCTTAAGATGTTGTGATATACCCCTGGGG 2212
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
739 IArgHisIleGlnIleGluAlaLeuGlnLysSerProHisArgSerVal 755
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2213 CTAGACATGTGTATATCCAAAGACGAGGCTTCTCTCATATTTCTTGT 2262
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGlySGL 772
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2263 ATTAAGAACCAGGCTACAGGCAATTATTAATTTAAATGGCAAGGGAGAGA 2312
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
772 uAlaThrSerArgThrPheThrAlaMetGlyLeuGluTyrGluAspAlaY 789
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2313 AGCCAAAGTCCGCGACCTTATAGATCTTGTGTGTGAGTGAATTAACA 2362
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
789 alGluAspAlaLysGlnSerLeuLysThrSergLysProLeuProGluAla 805
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2363 TTGAAGATGACATTGAAGCTTTCACACCGCATGGACCTTTACATATCT 2412
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
806 IleAlaIleLeuAlaLeuProProThrGluGlyGlyPyrArgSerSerg 822
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2413 GTTATGTGTTTGTATATA...COTCAAGAAATAGATACCCGCTTACGCT 2459
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
822 uAlaTyrLysTyrValIleHisGlnAspLeuLeuProLeuIleGlySera 839
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2460 GACATATTAAGTACATCATCATGAAAGACTCTGTACCTACATCAACACA 2509
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
839 snAsnValLeuLeuGluGluMetAspThrTyrGluTyrAlaLeuLysSer 855
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2510 ACAATGTCAATCCAGAAAGATTAAGTACTTTGAGATGGGCTTTGAAGAGC 2559
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
856 TrpAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTyr 872

```

```

|||||
2560 TGGTCTCAGTGTTCACAAACCCCTGTGGTGGAGTTCACAGTACCTAAATA 2609
872 TGLYSEGLUASPHISGLNHSGLYARGALATHRGlnLeuProGlyAla 889
2610 TGGATGCCGTAGAGAAAGTGATATAATAATGATCCATCGACGCTTGTGG 2659
889 SPHSLSYSASGPROLYSPROILEARGARGYCSASNGINHSIPRO 905
2660 AGGCCACCAAAAACCCCAACCTATTAGACGATGTCATATATCAAGAG 2709
906 CysSerGlnProValThrValThrGlnGlyLeuAlaCysSerArgse 922
2710 TGTACATCATCTCTGTGTAGAGAGAAATGGACACACCTCCCAAAAC 2759
922 TCGAGLGLYSGLNGLYALGlnThrArgGlyLeuGlnCysLeuLeuPro 939
2760 CTGTGGAAAGTCTGTGTATGAGCTTCGACACTGTCAGCTGCTTCAGGAC 2809
939 euserAsnglyThrHisLysValMetProAlaLysAlaCysAlaGlyASP 955
2810 TCTCTGATGGACCAACCCGCTCTGTGCACAGCAAAATCTGCATGGGTGAC 2859
956 ArgProGlnAlaArgArgProCysLeuArgValProCysProAlaGlnTr 972
2860 CGTCCCGAGACCCGCGCTGTACACAGATGCCCTGCGCTGCACAGCTG 2909
972 PArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyLeuGlyLeu 989
2910 GAACAACGAGACCCGTGAGTGAAGTGTTCAGTCACTCGCGTGAAGAACGG 2959
989 LnglnArgGlnValAlaCysArgThrAsnAlaAsnSerLeuGlyHisCys 1005
2960 AGGTGAGGACAGCTCTCTGACAGGCGTGGG.....GACCACTGT 2997
1006 GlnGlyAspArgProAspThrValAlaGlnValCysSerLeuProAlaCysG 1022
2998 GATGTTGAAGACCTCGAGTCGCTGACAGCCCTGTCACTGCTCTTGTAA 1047
1022 YGLYAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLeuGlyT 1039
3048 TGAAT..... 3051
1039 hrProGlnGlyGlnTrpValProGlnSerGlyProLeuHisProIleAsn 1055
3051 ..... 3051
1056 LysIleSerSerMetCysAlaAlaGlnProCysThrGlyAspArgSerVa 1072
3052 .....GACCAATGTTGGGAGACAAAGTCCAT 3077
1072 LPhecysGlnMetGluValLeuAspArgTyrCysSerIleProGlyTyrH 1089
8078 ATTCTGTCAATAGGAAGTGTGGCAGCATACGCTCCATACCAAGTTATA 3127
1089 LAsrLeuCysCysValSerCysIleLysLysAlaSerGly..... 1102
3128 ACAAGTTAATTGTGTGAGTCTTCACCAAGCCGAGTACAGACCCCTGCCACA 3177
1103 .....ProAsnPr 1105
3178 CCATACCTTCTAGAAAGTGTGTAACATCATGATCATCTCTAAACC 3227
1105 oGlyProAspProGly.....ProThrSerLeuProPrope. 1117
3228 TAGTGAACCTCCCTAGATCTCTAGTAGCGTACATCTTGTGTTCTTATAC 3277
1118 .....SerThrProGly.....SerProLeu 1124
3278 ATTTCAGAGACCCCTGCAAGAAGATGCTTTGAGTACGATCTCTTCAGAG 3327
1125 ProGlyProGlnAspProAlaAspAlaGlnProGlyLysProTh 1141
|||||

```

```

3328 GGAGGTCCAAATGCATATGCT...GCTTTCAGGCCCAACAGTAAACCTGA 3374
1141 TGLYSEGLUASPHISGLNHSGLYARGALATHRGlnLeuProGlyAla 1158
3375 TGGTCTCATTATACGCCAG.....AGGAGTCTCAGACAGACGAGAGTA 3418
1158 euAspThr.....SerSerProGlyThrGln 1166
3419 AGACTGTGAGACTGTGTACCCGATACCTCTCCCAACCCACAG 3462
seq_name: /SIND5/gcgdata/geneseq/geneseqn-emb1/NA1999.DMT:AAZ32056
seq_documentation block:
ID AAZ32056 standard; DNA: 5774 BP.
XX
AC AAZ32056;
XX
DT 10-JAN-2000 (first entry)
XX
DE Human METH2 related EST AB002364.
XX
KW Human; METH1, METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency; ss.
XX
OS Homo sapiens.
XX
PN W09937660-A1.
XX
PD 29-JUL-1999.
XX
PE 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
PR 28-AUG-1998; 98US-0098539.
XX
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX
PI IrueLa-Arispe L, Hastings GA, Ruben SM;
XX
WPI: 1999-590684/50.
XX
DR New isolated metalloprotease thrombospondin polypeptides, useful for
PT treating hyperproliferative disorders, cancers or autoimmune disorders
PT
PS Disclosure: Page 423-427; 457pp; English.
XX
AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC respectively. METH1 and METH2 have been found to be potent inhibitors of
CC angiogenesis both in vitro and in vivo. They can be used for treating
CC cancer and other disorders related to angiogenesis including abnormal
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC endometrial bleeding disorders, diabetic retinopathy, some forms of
CC macula degeneration, haemangiomas, and arterial-venous malformations.
CC They may be useful in treating deficiencies or disorders of the immune
CC system, by activating or inhibiting the proliferation, differentiation,
CC or mobilisation (chemotaxis) of immune cells. The etiology of these
CC immune deficiencies or disorders may be genetic, somatic, such as
CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC toxins), or infectious. They can also be used to treat inflammatory
CC conditions, both chronic and acute conditions. The products can also be
CC used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
CC AAY49511 represent sequences given in the exemplification of the present
CC invention.
XX
Sequence 5774 BP; 1766 A; 1129 C; 1258 G; 1621 T; 0 other;

```

alignment_scores:

Quality: 3615.50 Length: 1274
 Ratio: 3.956 Gaps: 22
 Percent Similarity: 71.743 Percent Identity: 53.375

alignment_block:

US-09-938-330-20 x AA232056 ..

Align seg 1/1 to: AA232056 from: 1 to: 5774

```

37 AsptYrGlyValThrValProCysSerThrAspPheArgGlyArgPheLeu 53
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
110 GAGTATGAGCTGCTGACTCCAGTACAGCAAAATCAGAAAGAGCGCTACT 159
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
53 uSerHisValValSerGlyProAlaAlaSerAlaGlySerMetValV 70
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
160 CTCCCACTACTCTTCT..... 175

70 aLAspThrProThrLeuProArgHisSerSerHisLeuArg..... 84
   ::::: ::::: ::::: ::::: ::::: ::::: :::::
176 ..... ValAlaArgSerProLeuHisProGlyGlyThrLeuTrpPr 196
   ::::: ::::: ::::: ::::: ::::: :::::
197 GCGAGGAGCGCTGCTCCAAACCT..... 220
   ::::: ::::: ::::: ::::: ::::: :::::
98 oGlyArgValGlyArgHisSerLeuThrPheAsnValThrValPheGlyL 115
   ::::: ::::: ::::: ::::: ::::: :::::
221 ..... GAGCACTGTCTTTAACTCACTCAGCGCATTTGGAA 254
   ::::: ::::: ::::: ::::: ::::: :::::
115 ySgLeuLeuHisLeuArgLeuArgProAsnArgArgLeuValValProGly 131
   ::::: ::::: ::::: ::::: ::::: :::::
255 AAGATTTCATCTCGAGCTAAAGCCCACTCAACTAGTACCTCTGGG 304
   ::::: ::::: ::::: ::::: ::::: :::::
132 SerSerValGluTrpGlnGlu..... 138
   ::::: ::::: ::::: ::::: ::::: :::::
305 GCTGTGTGGAGTGGCATGAGACATCTCTGTGCTGGGAATATAACCGCA 354
   ::::: ::::: ::::: ::::: ::::: :::::
139 ..... AspPheArgGluLeuPheA 145
   ::::: ::::: ::::: ::::: ::::: :::::
355 TCCCATTAACAACCATCAACCAAGAGCTACGTATATGAAATCCGGAATA 404
   ::::: ::::: ::::: ::::: ::::: :::::
145 rGlnProLeuArgGlnGlnCysValThrThrGlyValThrGlyMet 161
   ::::: ::::: ::::: ::::: ::::: :::::
405 CAGAGCCTTTGAGACTAACTGTCTTATGTGTGATCGATCGTGACATT 454
   ::::: ::::: ::::: ::::: ::::: :::::
162 ProGlyAlaAlaValAlaIleSerAsnCysAspGlyLeuAlaGlyLeu 178
   ::::: ::::: ::::: ::::: ::::: :::::
455 CCAGAACCTCTGTTGCCATCGACAACTGTCATGTGCTGCTGGAAATGAT 504
   ::::: ::::: ::::: ::::: ::::: :::::
178 eArgThrAspSerThrAspPhePheIleGluProLeuGluArgGlyGln 195
   ::::: ::::: ::::: ::::: ::::: :::::
505 AAAAGTATATATGAGAGTATTCATTGAACCTTGAAAGAGGTAAAC 554
   ::::: ::::: ::::: ::::: ::::: :::::
195 InGluValGluAlaSerGlyArgThrHisValValThrArgArgGluAla 211
   ::::: ::::: ::::: ::::: ::::: :::::
555 AGATGGAGGAGAAAGAAAGAGGATTCATGTGTCTACAAAGATCAGCT 604
   ::::: ::::: ::::: ::::: ::::: :::::
212 ValGlnGlnGluTrpAlaGluProAspGlyAspLeuHis..... As 225
   ::::: ::::: ::::: ::::: ::::: :::::
605 GTAGAACAGGCTCCCATAGACATTCACAAAGACTTCACACTACGACAGATC 654
   ::::: ::::: ::::: ::::: ::::: :::::
225 nGluAlaPheGlyLeuGlyAspLeuProAsnLeuLeuGlyValGlyVal 242
   ::::: ::::: ::::: ::::: ::::: :::::
655 GGACCTGGAGGCGCTGTATGATCTAGGTACTGTTTATGGCAACATCCACC 704
   ::::: ::::: ::::: ::::: ::::: :::::
242 spGlnLeuGlyAspThrGluArgLysArgArgHisAlaLysProGlySer 258
   ::::: ::::: ::::: ::::: ::::: :::::
705 AGCAGCTGAATGAAGAAACAATGAGAGCGCGCAGACCGGAGAGAAACGAT 754
   ::::: ::::: ::::: ::::: ::::: :::::
259 TyrSerIleGluValLeuLeuValValAspAspSerValValArgPheH 275

```

```

755 TACAAATATCGAGGTACTGCTGGAGTGGATGACTCTGTGCTCCGTTTCCA 804
   ::::: ::::: ::::: ::::: ::::: :::::
275 sGlyLysGluHisValGlnAsnThrValLeuThrLeuMetAsnIleValA 292
   ::::: ::::: ::::: ::::: ::::: :::::
805 TGCCAAAGAGCAGCTCCAAACTACTCTGACCTTAATGAACTATTGGA 854
   ::::: ::::: ::::: ::::: ::::: :::::
292 spGluIleThrHisAspGluSerLeuGlyValHisIleAsnIleAlaLeu 308
   ::::: ::::: ::::: ::::: ::::: :::::
855 ATGAAATTTACCATGATGAGTCCCTCGAGCTGCATATTAATGTGCTCGT 904
   ::::: ::::: ::::: ::::: ::::: :::::
309 ValArgLeuIleMetValGlyThrArgGlnSerLeuSerLeuIleGluAr 325
   ::::: ::::: ::::: ::::: ::::: :::::
905 GTCCGATATATATCTGGATATATCAAAAGTCCATCACCCTCATAGAAAG 954
   ::::: ::::: ::::: ::::: ::::: :::::
325 gGlyAsnProSerArgSerLeuGlnIleValCysArgTrpAlaHisSerG 342
   ::::: ::::: ::::: ::::: ::::: :::::
955 GGGAAACCATCCAGAGCTTGAGAAATGTGTGTGCTGGGCGTCCCAAC 1004
   ::::: ::::: ::::: ::::: ::::: :::::
342 InGlnArgGlnAspProSerHisAlaGluHisHisAspHisValAlaPhe 358
   ::::: ::::: ::::: ::::: ::::: :::::
1005 AGCAAGATCTGATCTCAACCACTCTGAACACATGACATGCAATTTT 1054
   ::::: ::::: ::::: ::::: ::::: :::::
359 LeuThrArgGlnAspPheGlyProSer..... GlyThrAlaProVa 372
   ::::: ::::: ::::: ::::: ::::: :::::
1055 TTRACCAAGCAAGACTTTGGACTCTGTGAATGCAAGGATTTGCTCAGT 1104
   ::::: ::::: ::::: ::::: ::::: :::::
372 ThrGlnMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAspG 389
   ::::: ::::: ::::: ::::: ::::: :::::
1105 CACCGCATGTGCATCCAGAGAGGATTTGACCTGATATATAGAGATG 1154
   ::::: ::::: ::::: ::::: ::::: :::::
389 LysPheSerSerLalaPheValIleAlaHisGluThrGlyHisValLeuGly 405
   ::::: ::::: ::::: ::::: ::::: :::::
1155 GTTTTTCATCTGCTTTTGTATAGCCATGAAGGCGCAATGCTTTGGGA 1204
   ::::: ::::: ::::: ::::: ::::: :::::
406 MetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuG 422
   ::::: ::::: ::::: ::::: ::::: :::::
1205 ATGAGACATGATGAGCAAGGCAAGGAGTGTGTATGAGACTGTAATGG 1254
   ::::: ::::: ::::: ::::: ::::: :::::
422 ySerValMetLalaProLeuValGlnAlaAlaPheHisArgPheHisTrpS 439
   ::::: ::::: ::::: ::::: ::::: :::::
1255 AAGTGTATGCTGCTCTTGTGATCAAGCAGCATTCATCGTTACCACTGGT 1304
   ::::: ::::: ::::: ::::: ::::: :::::
439 eArgCysSerLysLeuGlnLeuSerArgThrLeuProSerTyrAspCys 455
   ::::: ::::: ::::: ::::: ::::: :::::
1305 CCGGATGCACTGCTCAAGAACTGAAGATATATCCATTCTCTATGACTGT 1354
   ::::: ::::: ::::: ::::: ::::: :::::
456 LeuLeuAspAspProPheAspProAlaTrpProGlnProProGluLeuPr 472
   ::::: ::::: ::::: ::::: ::::: :::::
1355 CTCTTGTATGACCTTTTGTATCATGATTGGCCTTAACCTCCCAAACTTCC 1404
   ::::: ::::: ::::: ::::: ::::: :::::
472 oGlyTyrLeuSerThrSerMetAspGlnGlnCysArgPheAspPheGlySerG 489
   ::::: ::::: ::::: ::::: ::::: :::::
1405 TGGATTCATATTATTCTATGATGAGCAATGCTGTTTGTGATTTGGGTGG 1454
   ::::: ::::: ::::: ::::: ::::: :::::
489 LysThrGlnThrCysLeuAlaPheArgThrPheGlnProCysLysGlnLeu 505
   ::::: ::::: ::::: ::::: ::::: :::::
1455 GCTATTAAGATGWCACCGCGTTCGAACTTTTGAACCATTTAAACAGCTG 1504
   ::::: ::::: ::::: ::::: ::::: :::::
1505 TGGTGTGACCATCTCTGATTAATCCCTACTTTTGTAAAGACTAAAAAGGAGC 1554
   ::::: ::::: ::::: ::::: ::::: :::::
522 oProLeuAspGlyThrGlnCysAlaProGlyLysTrpCysPheLysGlyH 539
   ::::: ::::: ::::: ::::: ::::: :::::
1555 TCCACTTGATGGAGACTGAATGTGCTGTGAAATATGTCTTTAAAGGCTG 1604
   ::::: ::::: ::::: ::::: ::::: :::::
539 iCysIleTrpLysSerProGlnGlnThrTyrGlyGlnAspGlyGlyTrp 555
   ::::: ::::: ::::: ::::: ::::: :::::
1605 ATTGCATGTGAAGAAATGCTAATCAG...CAAAACAACATATGCAATTGG 1651
   ::::: ::::: ::::: ::::: ::::: :::::
556 SerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyValA 572
   ::::: ::::: ::::: ::::: ::::: :::::

```

```

1652 GGGTCATGACCTAAATTGGCTCTCTGTCGACATGTGGAACCTGCT 1701
572 LArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGly 589
1702 TCGTTTCAGAAACAGCAGTGCATTAATCCCATGTCATCATAGTGTCT 1751
589 rProCysLeuGlyProMetPheGluTyrGlnValCysAsnSerGlu 605
1752 AGGATTTGCCGGTGTAAATTTAGTACACGCTTTGTGTACACGAAGA 1801
606 CysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaTyr 622
1802 TGCACAAAACCTTTGAGAGACTTCAGACACAGAGTGCAGCAGGAA 1851
622 nSerTyrTyrValHisGlnAsnAlaLysHisSerTyrPvalPro 639
1852 CTCCACCTTGATATACAGATACCAAAACACCTGTTGCCATATGAC 1901
639 roAspAspAlaGlnLysCysGluLeuLeuLeuGlnSerAlaAsp 655
1902 ATCCGACCCCAAGAAAGATGCCACCTTACTGTCTGACGACGAGAC 1951
656 GLAspValValPheMetAsnGlnValValHisAspGlyThrArg 672
1952 GAGAGTGTGCTTACATGAACAACTGCTGATGTAACGACCTGTCT 2001
672 TTYrArgAspProTyrSerValCysAlaArgGlyGluCysValPro 689
2002 TTACAAATATCCATATGACATATGTGCGAGAGAGTGTGTAAGAG 2051
689 LysCysAspGluValGlySerMetLysAlaAspAspLysCysGly 705
2052 GCTGTGATTAAGAAATTTGTTCTTAATAGGTTGAGAGTAAGTGT 2101
706 CysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeu 722
2102 TGTGAGAGGATTAATTTCCACCTGCGAACCCTGAAGGACATTAC 2151
722 sAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnLeuPro 739
2152 AACTCCAGAGAGCTTGGGTACCTTAAGATGTTGATATACCCCTG 2201
739 LArgHisIleGlnIleGlnAlaLeuLysSerProHisArgSerVal 755
2202 CTAGACATGTGTTATCCAGAAAGAGAGGCTTCTCATATCTTCT 2251
756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGly 772
2252 ATTAGAACACAGCTACAGGCTTATATTTAAATGGCAAGGGAGGA 2301
772 uAlaThrSerArgThrPheThrAlaMetGlyLeuGlnTyrPgluAsp 789
2302 AGCCAAAGTGGGACCTTCATAGATCTTGTTGAGAGTGGATTTACA 2351
789 aGlnuAspAlaLysGlnSerLeuLysThrSerGlyProLeuProGlu 805
2352 TTGAAGATGACATTTGAAGCTTTCACACCGATGACCTTACATGAT 2401
806 IleAlaIleLeuAlaLeuProProThrGluGlyGlyProArgSer 822
2402 GTTATGTTTGTATATA...CTCAAGAAATGATACCCGCTGAGCT 2448
822 uAlaTyrLysTyrValIleHisGlnAspLeuProLeuIleGlySer 839
2449 GACATTAAGTACATCATCATGAAGACCTGTACCTACATCAACAGA 2498
839 snAsnValLeuLeuGluGlnMetAspThrTyrGluTyrPalaLeu 855
2499 ACAATGTCACTCAGAAAGATTAAGATCTTTGAGTGGCTTTGAAG 2548
856 TrpPalaProCysSerLysAlaCysGlyGlyGlyIleGlnPheThr 872
2549 TGCTTCAGAGTTCCAAACCTGTGTGTGAGGTTTCCAGTACACTRA 2598

```

```

872 rGlyCysArgArgArgAspHisIleMetValGlnArgHisLeuCys 889
2599 TGGATGCCCTAGCAAAAGATTAATTAATGTCATGACGCTTCTGTG 2648
889 sPHisLysLysArgProLysProIleArgArgArgCysAsnGlnHis 905
2649 AGGCCAACAAAGCCGAAACCTATTAGACGAAATGCAATTCATCAAG 2698
906 CysSerGlnProValTyrPalaThrGlnGluTyrPgluValCys 922
2699 TGTACACATCACTGCTGGGTAGCAGAGATGGAACTGCACCAAAAC 2748
922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuPro 939
2749 CTCTGCAATCTGCTGCTACACCTTGCGACCTGTACGCTGCTTAC 2798
939 euserAsnGlyThrHisLysValMetProAlaLysAlaCysAla 955
2799 TCCTTGATGACCAACACCTGCTGTGCACACGAAATACATGAGT 2848
956 ArgProGluAlaArgArgProCysLeuArgValProCysProAla 972
2849 CGTCCGAGAGCGCGCGCTGTAAACAGATGCCCTGCCCTGCACAG 2898
972 pArgLeuGlyAlaTyrSerGlnCysSerAlaThrCysGlyGluGly 989
2899 GAAGACAGAGCCCTGAGTGAAGTGTTCAGTGAACCTGCGTGAAGAC 2948
989 LnglnArgGlnValValCysArgThrAsnAlaAsnSerLeuGlyHis 1005
2949 AGCTGAGCGAGTCTCTGCGAGGCTGGG...GACCCTGT 2986
1006 GlnGlyAspArgProAspThrValGlnValCysSerLeuProAla 1022
2987 GATGGTGAAGAGCTGAGTGTGTCAGAGGCTGTCACTGCTGCTGTA 3036
1022 yGlyAsnHisGlnAsnSerThrValArgAlaAspValTyrPglu 1039
3037 TGAAT... 3040
1039 hrProGluGlyGlnTyrPvalProGlnSerClyProLeuHisPro 1055
3040 ... 3040
1056 LysIleSerSerMetCysAlaAlaGluProCysThrGlyAspArg 1072
3041 ...GAACTATGTTGGAGACAAAGTCCAT 3066
1072 LpneCysGlnMetGlnValLeuAspArgTyrCysSerIleProGly 1089
3067 ATTCTGCAATGAGAGTGTGGCAGCATCTCTCCATACAGGTTATA 3116
1089 sArgLeuCysCysValSerCysIleLysLysAlaSerGly... 1102
3117 ACAAGTTATGTGTGAGTCTCGACAGCAAGGACAGTACACCTGCC 3166
1103 ...ProAsnPr 1105
3167 CCATACCTTAGAAGCTGTGAAGTCAATGATGATCATCTTAACCC 3216
1105 oGlyProAspProGly...ProThrSerLeuProProPhe... 1117
3217 TAGGACCTCCCTGAGATCTTGTAGTATGCTCATCTTGTGCTTAT 3266
1118 ...SerThrProGly...SerProLeu 1124
3267 ATTACAGAGACCCCTGCAAGAGATGCTTTGAGTACATCTCTCAG 3316
1125 ProGlyProGlnAspProAlaAspAlaGluProProGlyLysPro 1141
3317 GGAGTCCAAATGATATGCT...GCTTTCAGGCAACAAAGTAAACCT 3363

```

```

1141 rglYserGIuASpHISglNHISglYArGAlarhrglNleuProglYAlaL 1158
|||||:|||||:|||||:|||||:|||||:
3364 TGGTGGTAATTTACGCCAG.....AGAGCTCTCAGCAAGAGAGATA 3407
|||||:|||||:|||||:|||||:|||||:
1158 euASpThr.....SerSerProglYThrGlnHISPro 1168
|||||:|||||:|||||:|||||:|||||:
3408 AGACTGTAGACTGTGTACCCATCCCTCCACCACCAAGAGGGTC 3457
|||||:|||||:|||||:|||||:|||||:
1169 .....PheAlaProgl 1172
|||||:|||||:|||||:|||||:|||||:
3458 CACCTAGTTCAGCTTCACAAATGGCTGCTGCTTCTTTGCAAGCCAG 3507
|||||:|||||:|||||:|||||:|||||:
1172 uThrProIleProglYAlaSerThrIleSerProThrThrProglYg 1189
|||||:|||||:|||||:|||||:|||||:
3508 TGATTCATAGGTGCTTCTTTCAGGCAAGAACCAAGAAAGATGCAA 3557
|||||:|||||:|||||:|||||:|||||:
1189 YLeu.....ProThrProglYThrGlnHISProThr... 1199
|||||:|||||:|||||:|||||:|||||:
3558 AGATCATTGACAAACAGACCTCG.....ACAAATCATCCACCTTA 3598
|||||:|||||:|||||:|||||:|||||:
1200 .....ProValPro.....GluAspLysglYgInProgl 1209
|||||:|||||:|||||:|||||:|||||:
3599 GAAAGATGAGAAAGTGAACCAAAAGGCTAGAAACAGAGAAACCTGG 3648
|||||:|||||:|||||:|||||:|||||:
1209 YGluAspLeuArGHisProglY 1216
|||||:|||||:|||||:|||||:|||||:
3649 ACAACCTCTCTCTCCATGGT 3670
|||||:|||||:|||||:|||||:|||||:

```

```

seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA95824
seq_documentation_block:
ID AAA95824 standard; cDNA; 5774 BP.
XX
AC AAA95824:
XX
DT 23-FEB-2001 (first entry)
XX
DE Human metalloproteinase KIAA0366 cDNA.
XX
KW Human: KIAA0366; ADAMTS; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vacuole; notochord; neuroprotective; antiparkinsonian;
KW cerebropoietic; cytoskeletal; antihypertensive; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury; ss.
XX

```

```

Homo sapiens.
Key Location/Qualifiers
CDS 2..3607
/*tag= a
/partial
/product= "KIAA0366"
FT
FT WO200053774-A2.
XX
XX 14-SEP-2000.
XX
XX PD 08-MAR-2000; 2000WO-US06237.
XX
XX PR 08-MAR-1999; 99US-0264585.
XX
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
XX PI Kerner GS, Clark M, Maki R;
XX
XX DR MPI; 2000-594326/56.
XX
XX DR P-PSDB; AAB21254.
XX
XX PT Polynucleotide encoding novel members of a disintegrin,
XX PT metalloproteinase and thrombospondin domain protein family used to
XX PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX

```

```

PS Example 3; Fig 7; 129pp; English.
XX
XX The present sequence encodes human metalloproteinase KIAA0366. KIAA0366
CC shows homology to the ADAMTS family of proteins, which contain
CC thrombospondin, integrin and metalloproteinase domains. ADAMTS
CC polypeptides are useful for the manufacture of medicaments for treating
CC conditions associated with neuroinflammation and/or neurodegeneration,
CC such as Alzheimer's disease, Parkinson's disease and stroke. They are
CC also useful for treating conditions associated with cell proliferation,
CC cell migration, inflammation and/or angiogenesis, such as cancer,
CC arthritis and autoimmune diseases. They can be used to treat patients
CC afflicted with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 5774 BP; 1766 A; 1129 C; 1258 G; 1621 T; 0 other;

```

```

alignment_scores:
Quality: 3615.50 Length: 1274
Ratio: 3.956 Gaps: 22
Percent Similarity: 71.743 Percent Identity: 53.375

```

```

alignment_block:
US-09-938-330-20 x AAA95824 ..

```

```

Align seg 1/1 to: AAA95824 from: 1 to: 5774

```

```

37 AspTYrGlyValThValProCysSerThrAspPheArGlyArPheLe 53
|||||:|||||:|||||:|||||:|||||:
110 GAGTATGAGCTGTGTGTCACGACGACCAAAATCTAGAACGACGCTATCT 159
|||||:|||||:|||||:|||||:|||||:
53 uSerHisValISerGlyProAlaIaIaSerAlaGlySerMetValY 70
|||||:|||||:|||||:|||||:|||||:
160 CTCCCATACTCTTCT..... 175
|||||:|||||:|||||:|||||:|||||:
70 aLAspThrProProThrLeuProArGHisSerSerHisLeuArG..... 84
|||||:|||||:|||||:|||||:|||||:
176 .....GCGAGTCAACAAAGAGGTCA 196
|||||:|||||:|||||:|||||:|||||:
85 .....ValAlaArGSerProLeuHISProglYgYThrLeuTrpP 98
|||||:|||||:|||||:|||||:|||||:
197 GCGAGGACGCTGTCTCCAACT..... 220
|||||:|||||:|||||:|||||:|||||:
98 OGlyArGValGlyArGHisSerLeuTYrPheAsnValThValPheGlyL 115
|||||:|||||:|||||:|||||:|||||:
221 .....GAGCAGTGTCTTTAACATCAGCGCATTTGGAA 254
|||||:|||||:|||||:|||||:|||||:
115 YSGluLeuHISLeuArGLeuArGProAsnArGArGLeuValProglY 131
|||||:|||||:|||||:|||||:|||||:
255 AAGATTTCATCTCGACATAAGCCCAACACCTCACTAGTAGCTCTGGG 304
|||||:|||||:|||||:|||||:|||||:
132 SerSerValGluTrpGlnGlu..... 138
|||||:|||||:|||||:|||||:|||||:
305 GCTGTGTGGAGTGGCATGACATCTCTGTGCTCCGGAATTAACGA 354
|||||:|||||:|||||:|||||:|||||:
139 .....AspPheArGLeuLeuPhe 145
|||||:|||||:|||||:|||||:|||||:
355 TCCCATTAACAAACCATCAACAGAAAGTCTAGCTATAGAAATCCGAAAA 404
|||||:|||||:|||||:|||||:|||||:
145 rglNProLeuArGInglucysValTYrThrGlyValThrGlyMet 161
|||||:|||||:|||||:|||||:|||||:
405 CAGAGCTTGTGACAGACTAGCTGTATGTTGTGTGACATCGGGACAT 454
|||||:|||||:|||||:|||||:|||||:
162 ProglYAlaIaValAlaIaIeSerAsnCySAspGlyLeuAlaGlyLeu 178
|||||:|||||:|||||:|||||:|||||:
455 CCAGAACCTCTGTGTGCCATCAGCAACTGTGATGTCTGCTGGAAATGAT 504
|||||:|||||:|||||:|||||:|||||:
178 eArGThrAspSerThrAspPheHeIlleGluProLeuGluArGlyGln 195
|||||:|||||:|||||:|||||:|||||:
505 AAAAAGTATATGAAGAGATTTTCATGACCCCTGGAAGAGGTAAAC 554
|||||:|||||:|||||:|||||:|||||:
195 IngluysGluAlaSerGlyArGThrHisValYAlTYrArGArGAla 211
|||||:|||||:|||||:|||||:|||||:
555 AGATGAGAGAAAGAAAGAGATTCATGTGTCTACAAAGATCACT 604
|||||:|||||:|||||:|||||:|||||:

```



```

2402 GTTATGTTGATTATA...CCTCAGAAATGATFACCCGCTCAGGCT 2448
822 uAATyTyrIyTyValIleHISglAspLeuProLeuIleGlySerA 839
2449 GACATATATGATATCATCCAGAAAGACTGTGATCCTACAAATCAACAGCA 2498
839 sPAsnValLeuLeuGluMetAspThrTyGluTrpAlaLeuIleSer 855
2499 ACAATGTCTCCAGAGAAATAGTACTTTGAGTGGGCTTGAAGAGC 2548
856 TrpAlaProCysSerLysAlaCysGlyIleGlnPheThrIyTy 872
2549 TGTGTCTCAGGTTCCAAACCTGTGTGAGAGTTCCAGTACCTAATA 2598
872 rGlyCysArgArgArgAspHisMetValGlnArgHisLeuGlyS 889
2599 TGGATGTCCCTAGAAAGATGATATAAATGTGTGAGTGTGAGTGTG 2648
889 sPHisLysLysArgProLysProIleArgArgCysAsnGlnHisPro 905
2649 AGGCCAACAAAAAGCCAAACCTATTAGACGATGTGCAATATTCAGAG 2698
906 CysSerGlnProValTrpValThrGluGluTrpGlyAlaCysSerA 922
2699 TGTACACATCCACTGTGGTAGCAAGAAATGTCACACCTGCACCAAA 2748
922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuPro 939
2749 CTGTGAGAGTTCTGTGCTATCAGCTTCGACAGTACGCTGCTTCA 2798
939 eUsrAsnGlyThrHisLysValMetProAlaLysAlaCysAlaGly 955
2799 TCCCTGATGGCAGCAACCGCTCTGCACAGCAAAATGCTGATGGG 2848
956 ArgProGluAlaArgArgProCysLeuArgValProCysProAlaGln 972
2849 CGTCCGAGAGACCGCCGCGCTGTACACAGAGTGCCTGCTGCACAG 2898
972 pArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyIleGly 989
2899 GAAAAACAGACCGCTGAGTAGTGTCTGACGCTGCGGTGAAGAACGG 2948
989 GlnGlnArgGlnValValCysArgGlnHisAsnLysSerLeuGlnHis 1005
2949 AGGTGAGGAGGCTCTCTGCGAGGCTGG...GACCACTGT 2986
1006 GluGlyAspArgProAspThrValGlnValCysSerLeuProAlaG 1022
2987 GATGTGAAAAAGCCTGAGTGGGTGAGAGCTGTCAACTGCTCTCTGTA 3036
1022 yGlyAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLeu 1039
3037 TGAT..... 3040
1039 hrProGluGlyGlnTrpValProGlnSerGlyProLeuHisProIle 1055
3040 ..... 3040
1056 LysIleSerSerMetCysAlaAlaGluProCysThrGlyAspArgSer 1072
3041 .....GAAACCATGTTTGGGAGACAAAGCCAT 3066
1072 LpHeCysGlnMetGluValLeuAspArgTyrcysSerIleProGlyTy 1089
3067 ATTTCTGCAATGGAAGTGTGGCAGCATGCTCCATACAGGTTATA 3116
1089 IsArgLeuCysCysValSerCysIleLysLysAlaSerGly..... 1102
3117 ACAAGTTATTTGTGAGTCTCTGACGACGAGTAGCACCTGTGCACCA 3166
1103 .....ProAsnPr 1105

```

```

3167 CCATACCTTCTAGAAAGCTGTGAACATCATGATGATCATCTTAACCC 3216
1105 oGlyProAspProGly.....ProThrSerLeuProPhe..... 1117
3217 TGTGACCTCCCTAGATATCTCTAGTATGCTCTACATCTTTGGTCTTAT 3266
1118 .....SerThrProGly.....SerProLeu 1124
3267 ATTCAGAGACCCCTCAAAAGATGTCTTGTAGTAGCATCTCTTCACTG 3316
1125 ProGlyProGlnAspProAlaAspAlaAlaGluProGlyLysProTh 1141
3317 GGAGGTCCAAATGCATATGCT...GCTTTCAGGCCAAACGTAAACCTGA 3363
1141 rGlySerGlnAspHisGlnHisGlyArgAlaThrGlnLeuProGlyAla 1158
3364 TGTCTCATATTTACGCGAG.....AGGAGTGTCTCAGCAGCAGAGGA 3407
1158 eUsrPThr.....SerSerProGlyThrGlnHisPro 1168
3408 AGACTGTGAGACTGTGTACCGTACCATCTCTCCACCCACCAAGAGGTC 3457
1169 .....PheAlaProGly 1172
3458 CACCTCAGTTCAGCTTCACAAATGGCTGCTGCTCTCTTTCAGCGCAG 3507
1172 uThrProIleProGlyAlaSerTrpSerIleSerProThrTrpProGly 1189
3508 TGATTCAATAGTGTCTCTCTCTCAGCAGAACCTCAAAAGATGGA 3557
1189 LysLeu.....ProThrPlyThrPheGlnThrProThr... 1199
3558 AGATCATTTGACAAACAGACGCTCG.....ACAAATATATCCACTTA 3598
1200 .....ProValPro.....GluAspLysGlyGlnProGly 1209
3599 GAAAGATGAGAAAGTGAACCAAAAGCTAGAAACAGAGAGAAACCTGG 3648
1209 yGluAspLeuArgHisProGly 1216
3649 ACAACCTCTCTCTTCCATGCT 3670

seq_name: /SID5/gc9data/geneseq/geneseqn-emb1/NA2001A.DAT:AAAC90313
seq_documentation_block:
ID AAC90313 standard; DNA; 5774 BP.
AC AAC90313;
DT 19-MAR-2001 (first entry)
XX
DE AB002364 cDNA clone.
XX
XX METH; metalloproteinase; thrombospondin; angiogenesis inhibition;
XX cancer therapy; benign tumour; ocular angiogenic disease;
XX rheumatoid arthritis; psoriasis; wound healing; endometriosis;
XX vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
XX scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
XX coronary collateral; cerebral collateral; arteriovenous malformation;
XX ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
XX plaque neovascularisation; telangiectasia; haemophilic joint; EST;
XX angiofibroma; fibromuscular dysplasia; expressed sequence tag;
XX Crohn's disease; atherosclerosis; birth control; ss.
XX
OS unidentified.
XX
XX PM WO200071577-A1.
XX
XX PD 30-NOV-2000.
XX
XX PF 25-MAY-2000; 2000WO-US14462.
XX
XX PR 25-MAY-1999; 99US-0318208.

```



```

1105 CACGGCATGTGTATCATCATGAGAAATTGTACCTGCAATCATGAGATG 1154
389 lpheserSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 405
1155 GTTTTTCATCTGCTTTGTAGTAGCCATGAACGGCCATGTGTGGGA 1204
406 MetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuGly 422
1205 ATGACACATGATGAGAACAGGCAAGGTGTGTGATGAGACTGCTATGGG 1254
422 yserValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpS 439
1255 AAGTGCATAGCTCCCTTGTGTACAAGCAGCATTCATCTTACACATGCT 1304
439 eArgCysSerLysLeuGluSerArgTyrLeuProSerTyrAspCys 455
1305 CCCGATGCAGTGTCAAGAACTGAAAGATATATCCATTCCTTAAGACTG 1354
456 LeuLeuAspAspProPheAspProAlaTrpProGlnProGluLeuPr 472
1355 CTCCTTGATGACCTTTTGATCATGATGTGGCTTAACCTCCAGAACTTC 1404
472 oGlyLysAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerG 489
1405 TGGAAATCAATTATCTATGATGAGCAATGTCTTTGATTTTGTGTG 1454
489 lYTyGlnThrCysLeuAlaPheArgThrPheGluProCysGlyGlnLeu 505
1455 GGTATAAATGTGCAACCGCTTCCGAACTTTGACCCATGTAACACACTG 1504
506 TrpCysSerHisProAspAsnProTyrPheCysLysThrLysGlyLys 522
1505 TGGTGTACCATCTCGATTAATCCCTACTTTGTGAAGCTAAAGGAGACC 1554
522 oProLeuAspGlyThrGluCysAlaProGlyLysTyrCysPheLysGly 539
1555 TCCACTTATGAGCATGATGTGTCTGCTGGAATAATGTCTATAAGGCTC 1604
539 lscYailerTrpLysSerProGluGlnThrTyrGlyGlnAspGlyGlyTrp 555
1605 ATTCGATGTGGAGAAATGCTAATCAG...CAAAACAAGATGGCAATGG 1651
556 SerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyLys 572
1652 GGGTATGACTAAATTTGGCTCTGCTTCGAGACATGTGAAGTGTGTG 1701
572 lArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGlyGly 589
1702 TCGTTTCAGAACACAGCCAGTGCATTAATCCCATGCCCATCAATGTGTC 1751
589 rArgProCysLeuGlyProMetPheGlnTyrGlnValCysAsnSerGluGln 605
1752 AGGATGTCTGTGTTAATTTTGTAGTACAGACTTTGTAAACAGAGAA 1801
606 CysTrpGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgAs 622
1802 TGCCAAAACACTTTGAGACTTCAAGACACACAGCTGTGACAGAGAA 1851
622 nSerTyrTyrValHisGlnAsnAlaLysHisSerTrpValProTyrGluP 639
1852 CTCGCATTTGAAATACAGAAATACCAACACCACTGTTGCCATATGAAC 1901
639 roAspAspAlaGlnLysCysGluLeuLeuGlnCysGlnSerAlaAspThr 655
1902 ATCCGTGACCCCAAGAAAGATGCCACCTTTACTGTACGTCCAGAGACT 1951
656 GlyAspValValPheMetAsnGlnValHisAspGlyThrArgCysSe 672
1952 GGAGATGTTCTTACATGAACAACAGTGTGATGATGAGAACGCACTGTT 2001
672 rTyArgAspProTyrSerValCysAlaArgGlyGlnCysValProValG 689
2002 TTACAAAGATCCATATGATGTGTGCGAGAGAGTGTGTGAAGTGG 2051

```

```

689 lCysAspLysGluValAlGlySerMetLysAlaAspAspLysCysGlyVal 705
2052 GCTGTATGAAGAAATGTGTTCTATTAAGTTGAGGATTAAGTGTGTGTC 2101
706 CysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGlyLys 722
2102 TGTGGAGGATTAATTCACACTGCCGAACCGTGAAGGGGACATTTACAG 2151
722 sAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnLeuProAlaGly 739
2152 AACTCCAGAGAGCTGGGTACTTAAGATCTTTGATATACCCCTGGGG 2201
739 lArgHisIleGlnIleGlnAlaLeuGluLysSerProHisArgSerVal 755
2202 CTAGACATGTGTATTCAGAGAAAGAGGCTTCTCCATATTCCTGCT 2251
756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGly 772
2252 ATTAAGAACACAGGCTACAGGCCATTAATTTAAATGGCAAGGGGAGGA 2301
772 uAlaThrSerArgThrPheThrAlaMetGlyLeuGlnTrpGluAspAla 789
2302 ACCCAAGTCGGGACCTTCATAGATCTTGTGTGAGTGGGATTAATACA 2351
789 aGluAspAlaLysGluSerLeuLysThrSerGlyProLeuProGluAla 805
2352 TTGAAGATGACATTGAAGTCTTCACACCGATGACCTTATCATGATCCT 2401
806 lIleAlaIleLeuAlaLeuProProThrGlnGlyGlyProArgSerIle 822
2402 GTTATGTTTGTGATTA...CTCAAGAAAATGATACCGCTGTAGCCT 2448
822 uAlaTyrLysTyrValIleHisGluAspLeuLeuProLeuIleGlySer 839
2449 GACATATGAATACATCATCATGAGACCTGTGATACCTACATCAACAGCA 2498
839 sPheValLeuLeuGlnGluMetAspThrTyrGluThrAlaLeuLysSer 855
2499 ACAATGTATCCAGAAAGATGATTAATGATCTTTGAGTGGGCTTTGAAGAG 2548
856 TrpAlaProCysSerLysAlaCysGlyGlyGlyIleGlnPheThrLysTy 872
2549 TGTGTCTCAGGTTTCCAAACCTGTGTGGAGGTTTCCAGTACCTTAATA 2598
872 rGlyCysArgArgArgArgAspHisHisMetValGlnArgHisLeuCys 889
2599 TGGATGCCGTAGGAAAGATGATTAATAATGTGCATTCGACGCTGTG 2648
889 sPheLysLysArgProLysProIleArgArgArgCysAsnGlnHisPro 905
2649 AGGCCAACAAAGAACCGAAACCTATTAGAGAAATGTCAATTTCAAGAG 2698
906 CysSerGlnProValTrpValThrGluGluTrpGlyAlaCysSerArgSe 922
2699 TGTACATCATCCTCTGGGTAGCAGAAAGATGGAGAACACTGCACCAAA 2748
922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysIleLeuProL 939
2749 CTGTGGAAGTTCTGGCTATCAGCTTGCACCTGTACCTCTTACAGCAC 2798
939 eHisSerGlnTyrHisLysValMetProAlaLysAlaCysAlaLysAsp 955
2799 TCCTTATGACCAACACCGCTGTGTACACAGCAATACTGTGGGTAGC 2848
956 ArgProGluAlaArgArgProCysLeuArgValProCysProAlaGlnTr 972
2849 CGTCCGAGAGCGCGCGCTGTGAACAGAGTGCCTGCCCTGCACAGT 2898
972 pArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyGlnGlyIleG 989
2899 GAAACAGACACCTGTGAGTGTGAGTGTTCAGTACCTGTGTAAGAGAACG 2948

```

[illegible]

```
seq_documentation_block:
ID   AAV06594 standard; cDNA; 4580 BP
```

AC AAV06594 ;

DT 06-JUL-1998 (first entry)

Bovine N-proteinase cDNA.

fibrosis: Fh]ers-Dan]os disease: diagnosis; gene therapy; ss.

XX
OS Bos taurus

XX	Key	Location/Qualifiers
FH		

```

CDS      / *tag= a
FF       .
FF       .

```

```

/chain1_except= (pos:756..758, aa:Phe)

```

```

111         /transl_except= (pos:1770..1772, aa:Thr)
112
113     FT

```

```

FTT      /transl_except= (pos:2565--2567, aa:Thr)

```

PN W09800555-A1

PD 08-JAN-1998.

03-JUL-1997; 97WO-US12427

PR	02-JUL-1997;	97US-0886333
PR	02-JUL-1996;	96US-0021203

(CONT'D) COTTAGE A
XX
XX
P2

PA (LAPI/) LAPIERE C.
PA (PROC/) PROCKOP D. J.

XX	colide A,	Prockop DJ;
PI	Lapiere C,	

AA
DR
WPI: 1998-086980/08

DK FDDP, 1979/0000
XX

produce mature collagen in vitro and antibodies to treat fibrosis

XX

[illegible]

CC N-proteinase (see AAW47030), the enzyme responsible for cleaving

full-length cDNA was obtained by constructing a bovine cDNA

CC based on partial amino acid sequence of N-proteinase, and

CC with longer degenerate oligonucleotides (see AAV06595 and AAV06596) and that the oligonucleotide can be utilised in the recombinant

CC production of bovine N-proteinases in claimed transformed or

CC human N-proteinase clones (see AAV06392-93).

Sequence 4580 BP; 959 A; 1365 C; 1370 G; 800 T, 0 others, 5Q

alignment scores:

Quality:	3330.00	Deng.cn:	1230
Ratio:	3.729	Gaps:	24

Percent Similarity: 11.003 Percent Accuracy: 99.999

```

alignment_block:
ITS-09-938-330-20 X AAV06594 ..

```


730 GGCGTCTAGAGAGACAGCCACAGCTCGAGGGGAGGAGGACGACGCA 779
253 SALALysProGlySerTyrSerIleGluValLeuValAlaAspAsp 270
780 TGCTGCACACATGATCAATCAATCGAGGTCCTGCGCGCTGGATGACT 829
270 eValValArgPheHisGlyLysGluHisValGlnAsnTyrValLeuThr 286
830 CTGTGTCAGATGCCACGGGAAGAGCACGTACAGAAATACCTCTGACA 879
287 LeuMetAsnIleValAspGluIleTyrHisAspGluSerLeuGluValHis 303
880 CTCATGAACATGTCAATGAATACCATATGACATGACATGCTTGGTCCCA 929
303 sIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrArgInsert 320
930 CATCAACGTGGTCTCGTGGGATCATCTCTGATGTAAGAAATGCA 979
320 euSerLeuIleGluArgGlyAsnProSerArgSerLeuGluGlnValCys 336
980 TGAGCCATATCGAGATCGGAACCCCTCTCAGAGCTGGAGAAATGTCGC 1029
337 ArgTPrAlaHisSerGlnArgGlnAspProSerHisAlaGluHisH 353
1030 CGCTGGGCTACCTCCAGCAGAGACACAGCGGCCACGATGAAATACA 1079
353 sAspHisValAlaPheLeuThrArgGlnAspPheGlyProSer..... 367
1080 CGATCAGCGCATCTTCCTCACACGCGAGACTTTGGCCTTCGCGCATGC 1129
368 ..GlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAla 383
1130 AAGGTATGCTCTCTTCACCGGATGTCATCGGTCGCGACTGCGACC 1179
384 LeuAsnHisGluAspGlyPheSerSerAlaPheValIleAlaHisGluTh 400
1180 CTGAACCATAGAGAGCGGCTTCTCTCAGAGTTTGTGGCCCATGAGAC 1229
400 rGlyHisValLeuGlyMetGluHisAspGlyGlnLysnGlyCysAla 417
1230 TGGCCACGTGCTGGGATGAGACACGAGCGGAGGCAACCGCTGGCG 1279
417 spGluThrSerLeuGlySerValMetAlaProLeuValGlnAlaIlePhe 433
1280 ACGAGTGGCGGCTGGGACATCATGCGCCCTGCTGGAGCGCCCTTC 1329
434 HisArgPheHisTyrSerArgCysSerLysLeuGluLeuSerArgTyrLe 450
1330 CACCGCTTCACGTGCTCCGCTGACGCCACGAGAGCTGAGCCGCTACT 1379
450 upProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTyrProG 467
1380 GCACCTCTTACGCTGCTGCTGATGACCCCTTCCGCCACGACTGGCGG 1429
467 InProProGluLeuProGlyIleAsnTyrSerMetAspGluGlnCysArg 483
1430 CGCTGCCACAGCTCCGGAGACTGACATCTCATGAACGAGCAATGCCGC 1479
484 PheAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheG 500
1480 TTTGACTTGGCTGGCTGATCATGATGACGCGCTTCGGGCTTTTGA 1529
500 upProCysLysGlnLeuThrCysSerHisProAspAsnProTyrPheCysL 517
1530 CCCCTCAAGACATGCTGGTGGAGCATCTGTGACAAACCCCTACTTTTGA 1579
517 ysrThrLysLysGlyProProLeuAspGlyThrGluCysAlaProGlyLys 533
1580 AGACCAAGAGAGGGGCCCTTGGAGCGGACTATGTGTCACCTGGCAAG 1629
534 TrrCysPheLysGlyHisCysIleTrrPlySerProGluGlnThrTrrG 550
1630 CATTTGTTTAAAGACATGATCTGGCTGACACTTGAC...ATCTCAA 1676

550 yGlnAspGlyGlyTrrPserSerTrrThrLysPheGlySerSerArg 567
1677 ACGGAGGCGACATGGGGCGCTTGAGTCCGTTGGCTCTCTACGTA 1726
567 ercCysGlyGlyValAlaArgSerArgSerArgSerCysAsnAsnProSer 583
1727 CCTGTGGACAGGGCGCTGAAGTTCAGGACCGCCAGTGTGACAACCCACAC 1776
584 ProAlaTrrGlyArgProCysLeuGlyProMetPheGluTyrGlnVal 600
1777 CCGGCCAAGGGGGCCGACCTGCTGGCCCTTGCCTGACACTTCAGCT 1826
600 LysAsnSerGluGluCysProGlyThrTrrGluAspPheAlaGln 617
1827 CTCAGCGCCGACAGACTGCCCGACCTCCCTGGCTGACTTCCCGAGGAC 1876
617 IncCysAlaLysArgAsnSerTrrTyrValHisGlnAsnAlaLysHisSer 633
1877 AGTCCGCGCAGTGGACCTGTACTTCAGACACGCGACGCCACAC 1926
634 TrrValProTrrGluProAspAspAlaGlnLysCysGluLeuIleCys 650
1927 TGGCTGCCACAGACACCGGATGCCAAGAGAGATGCCACTGTACTG 1976
650 sGlnSerAlaAspThrGlyAspValAlaPheMetAsnGlnValHisAla 667
1977 CGAGTCCAGGAGACACCGGAGGTGCTGTCCATGAACCGCATGTCATG 2026
667 spGlyThrArgCysSerTrrArgAspProTrrSerValCysAlaAlaGly 683
2027 ATGGAGCGCGCTCTCTTCACAGCGCCCTTCAAGCTGTGTGGCGGG 2076
684 GluCysValProValGlyCysAspLysGluValGlySerMetLysAlaAs 700
2077 GACTGCAGAGAGTGGCTGTGACGCTGTATGCGCTCCAGCAAGACAGA 2126
700 pasPlyCysGlyValCysGlyLysAspAsnSerHisCysArgThrVal 717
2127 AGACAAAGTGGCGTGTGCGAGGGGACAAACGCACTGCAAGGTGCTCA 2176
717 ysrGlyThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuVal 733
2177 AGGGCAGTTCACACGCTCACCCCAAGAGCATGTTTACATCAAGATGTT 2226
734 GlnIleProAlaGlyAlaArgHisIleGlnIleGluAlaLeuGluLysSe 750
2227 GAGATCCCTGAGAGGACGACACCTGCTCATGAGAGGTAGACGCCAC 2276
750 rProHisArgSerValValLysAsnGlnValThrGlySerPheIleLeu 767
2277 CAGCCACCATGTCGCGCTCAAGAACCTGAGACAGGCAAGTCACTTCA 2326
767 snProLysGlyLys...GluAlaThrSerArgThrPheThrAlaMetGly 782
2327 ATGAAGACAATGAGTGTGATGCCAGTTCCAAACCTTCATGGCATGGCG 2376
783 LeuGluTrrGluAspAlaValGluAspAlaLysGluSerLeuLysThrSe 799
2377 GTGAGTGGCGATACAGAGACAGAGACGGCGGAGACGCTGAGACAT 2426
799 rGlyProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGlu 816
2427 GGGGCCCTCCAGCGCACATACCGCTTGTGGTATGCCG....GTGG 2470
816 lylGlyProArgSerSerLeuAlaTrrLysTrrValIleHisGluAspLeu 832
2471 GAGACACCGGGGTCTCACTGACGTACAAATACATGATTCATGAGACTCA 2520
833 LeuProLeuIleGlySerAsnAsnValLeuLeuGluGluMetAspThrTrr 849
2521 CTGAAT...GTGATGACACACAGCTCTGTGAAGAGGACTGTGTGTCTA 2567

```

849  rglutPAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGlyG 866
      |||||.....:|||||.....:|||||.....:
2568  CGAGTGGGCCCTGAGAGAGTGTCTCCGTCTCAAGCCCTGGGCGAG 2617
      |||||.....:|||||.....:|||||.....:
866  lylleGlnPheThrLysTyrGlyCysArgArgArgArgPheHisMet 882
      |||||.....:|||||.....:|||||.....:
2618  GGTCGCCAGTTCACCAAGATATGCTGCCGCGAGGCTGGACCAAGATG 2667
      |||||.....:|||||.....:|||||.....:
883  ValGlnArgHisLeuCysAspHisLysLysArgProLysProLLeuArg 899
      |||||.....:|||||.....:|||||.....:
2668  GTACACCGCTGGCTTGTGTGGCCCTCTCGAACCCCAAGCCATCGCGAG 2717
      |||||.....:|||||.....:|||||.....:
899  gArgCysAsnGlnHisProCysSerGlnProValTrpValThrGluGlu 916
      |||||.....:|||||.....:|||||.....:
2718  AGCGTGCACACCCAGAGAAATGCTCCAGCAGTGGGTGACAGCGGAT 2767
      |||||.....:|||||.....:|||||.....:
916  rpgLysAlaCysSerArgSerCysGlyLysLeuGlyValGlnThrArgly 932
      |||||.....:|||||.....:|||||.....:
2768  GGGAGCCATGTAGCCAGACCTGTGGCGGAGCAGCATGACGCTGGCTCC 2817
      |||||.....:|||||.....:|||||.....:
933  lIleGlnCysLeuLeuProLeuSerAsnGlyThrHisLysValMetProAl 949
      |||||.....:|||||.....:|||||.....:
2818  GTGCGCTGCATTACGCCGTACACGACACACCCCGCTCCGTCCAGC 2867
      |||||.....:|||||.....:|||||.....:
949  alYsAlaCysAlaGlyAspArgProGluAlaArgArgProCysLeuArgV 966
      |||||.....:|||||.....:|||||.....:
2868  CAAGCACTGCATGACGCCGCCGCGAGAGCCGCCGCTCCAGCCCG 2917
      |||||.....:|||||.....:|||||.....:
966  alProCysProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAla 982
      |||||.....:|||||.....:|||||.....:
2918  AGCTCGCCCTGGTCTGTGGCAGCCGCGCTGTGCTCCAGTGTCACTA 2967
      |||||.....:|||||.....:|||||.....:
983  ThrCysGlyGluGlylIleGlnGlnArgGlnValCysArgThrAspAl 999
      |||||.....:|||||.....:|||||.....:
2968  ACCTGTGGCAGCAGCAGCAGGAGCGCCAGTGCCTGCCAGCCCGCGA 3017
      |||||.....:|||||.....:|||||.....:
999  aAsnSerLeuGlyHisCysGlyGlyAspArgProAspThrValGlnValC 1016
      |||||.....:|||||.....:|||||.....:
3018  CGACACTCTGGCATCTGCCAGAGAGCGTCTGAGACAGCAGAGACT 3067
      |||||.....:|||||.....:|||||.....:
1016  ySerLeuProAlaCysGlyGlyAsnHisGlnAsnSerThrValArgAla 1032
      |||||.....:|||||.....:|||||.....:
3068  GCAGGCTTGCCCTGTCCCGCAACATCTAGATCCCTCCAGAGAGAGC 3117
      |||||.....:|||||.....:|||||.....:
1033  AspValITrpgLnuGlyThrProGluGlyLntTrpAlaProGlnSerG 1049
      |||||.....:|||||.....:|||||.....:
3118  TACGTA.....:GTCAGTGGCTGTCCGCCCGGA 3146
      |||||.....:|||||.....:|||||.....:
1049  yProLeuHisProIleAsnLysIleSerSerMetCysAlaAlaGluProC 1066
      |||||.....:|||||.....:|||||.....:
3147  CCCCAGCTCGCCCATCCGGAAGATCTCTCA.....:AAGGCCACT 3187
      |||||.....:|||||.....:|||||.....:
1066  ySThGlyAspArgSerValPheCysGlnMetGluValLeuAspArgTyr 1082
      |||||.....:|||||.....:|||||.....:
3188  GCCAAGCGCACAAGTATTTCTGTAGAGTGAAGTCTTCTCCGCTAT 3237
      |||||.....:|||||.....:|||||.....:
1083  CysSerIleProGlyTyrHisArgLeuCysCysValSerCys..... 1096
      |||||.....:|||||.....:|||||.....:
3238  TGCTCATCTCCAGGCTACACACACTGTCTCTCAAGTCTGTAACTGT 3287
      |||||.....:|||||.....:|||||.....:
1097  .....lIeLysLysAlaSerGlyProAspProGlyPro. 1107
      |||||.....:|||||.....:|||||.....:
3288  CACACACTCACCACAGTGGAGGAGCAGATAGACCCCTGGGGAAGC 3337
      |||||.....:|||||.....:|||||.....:
1108  .....AspProGlyProThrSerLeuProProPheSerThrPro 1120
      |||||.....:|||||.....:|||||.....:
3338  ACAACGACATTGACGTGTATGCTTACCTCCAGTGGCCACTGTAGGC 3387
      |||||.....:|||||.....:|||||.....:
1121  GlySerProLeuProGlyProGlnAspProAlaAspAlaAlaGluProP 1137
      |||||.....:|||||.....:|||||.....:
3388  ATGGAGGCGGCGCATACCAAGACCCCTGAGGTCCCTCTCAATGC 3437
      |||||.....:|||||.....:|||||.....:
1137  oGlyLysProThrGlySerGluAspHisGlnHisGlyArgAlaThrGlnL 1154

```

```

seq_name: /SID57/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV06592
seq_documentation_block:
ID  AAV06592 standard; cDNA; 6692 BP.
XX
XX  AAV06592;
AC
XX  06-JUL-1998 (first entry)
XX
XX  Human N-proteinase (long form) cDNA.
DE
XX
XX  N-proteinase; human; collagen; antibody; rheumatoid arthritis;
KW  fibrosis; Ehlers-Danlos disease; diagnosis; gene therapy; ss.
XX
XX  Homo sapiens.
OS
XX
XX  Key      Location/Qualifiers
FH  CDS      82..3717
FT  CDS      /*tag= a
XX
XX  WO9800555-A1.
XX
XX  08-JAN-1998.
XX
XX  03-JUL-1997; 97MO-US12427.
XX
XX  02-JUL-1997; 97US-0886333.
XX  03-JUL-1996; 96US-0021203.
XX
XX  (COL1) COL1GE A.
XX  (LAP1) LAP1ERE C.
XX  (PROC) PROCKOP D J.
XX
XX  Colige A, Lapiere C, Prockop DJ;
XX
XX  WPI; 1998-086980/08.
XX  P-PSDB; AAM47028.
XX
XX  Polynucleotide sequence encoding human N-proteinase - used to
XX  produce mature collagen in vitro and antibodies to treat fibrosis
XX  and rheumatoid arthritis
XX
XX  Claim 1; Fig 1A-C; 49p; English.
XX
XX  This nucleic acid sequence includes a coding region for the 130 kDa
XX  long form (see AAM47028) of human N-proteinase, the enzyme
XX  responsible for cleaving N-propeptide from procollagen to produce
XX  mature collagen. Overlapping cDNA clones were isolated from a
XX  human skin fibroblast cDNA library using bovine N-proteinase cDNA
XX  (see AAV06594) as probe, and used to generate the full-length cDNA.
XX  A nucleic acid sequence (see AAV06593) for the 70 kDa short form (see
XX  AAM47029) of human N-proteinase was also generated. The isolated
XX  human N-proteinases in claimed transformed or transfected host cells.
XX  They can also be used diagnostically for detection of N-proteinase
XX  expression and in gene therapy approaches to remedy deficiencies
XX  of functional N-proteinase in various disease states such as
XX  Ehlers-Danlos disease. The N-proteinase may be used for the
XX  production of mature collagen in vitro and for the production of
XX  antibodies which may be used for diagnosis and therapy of diseases
XX  including fibrosis and rheumatoid arthritis.

```

SQ Sequence 6692 BP; 1469 A; 1925 C; 1896 G; 1402 T; 0 other;

```
alignment_scores:      Length: 1207  
                        Quality: 3299.50  
                        Ratio: 3.707  
Percent Similarity:    Gaps: 20  
                       Percent Identity: 52.610
```

```
alignment_block:
US-09-938-330-20 x AAV06592 . .
```

Align seg 1/1 to: AAV06592 from: 1 to: 6692

[illegible]

238	glyLeuValAlaAspGlnLeuGlyAspTrpGlnArgLys...	ArgArgTrp	253
1111	1111
811	GCGCTCTAGAGAGACCGCCAAACAGCTCGAGGGAGGACGACGACG		860
253	salAlaLysProGlySerTyrSerIleGluValLeuLeuValAspAspS		270
1111	1111
861	TGCTGCACAGAGATGACATCAAAATCGAGGCTCTGGGGGTGATGACT		910
270	erValValArgPheHisGlyLysGlnHisValGlnAsnTyrValLeuTrp		286
911	CTGTGGTGCAGATTCCACGGGAGAGGACGATACAGAAGTACCTGGTGA		960
287	LeuMetAsnIleValAspGlnIleTyrHisAspGlnSerLeuGlyValHis		303
961	CTCATGACATTCCTGCATGGAATGTACCATGACGAGATCCTTGGGTGCCA		1010
303	siLeaSnIleAlaLeuValArgLeuIleMetValGlyTyrArgGlnSerI		320
1011	CATCAACGTGTCTCTGGTGGGATCATCTCTGATGATAGGAAGTCCA		1060
320	euseTrpIleGluArgGlyAsnProSerArgSerLeuGlnValCys		336
1061	TGAGCTCATGAGATGGGAAACCCCTCTCAGAACCTGGGAATGTGTGC		1110
337	ArgTrpAlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHisI		353
1111	GCGTGGGCTCACTCCACAGCAAAACGACACGGGCCACAGTGAATACCA		1160
353	asPheHisValAlaPheLeuTrpArgGlnAspPheGlyProSer.....		367
1161	CGATACGCCCATCTTCCCTACACAGGACGACTTTGGGCTTCGGGATGC		1210
368	..GlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAla		383
1211	AGGCTATGCTCTCTGTCCACGGCATGTGCTCATCTCGGTTCCGAGTGAAC		1260
384	LeuAsnHisGlnAspGlyPheSerSerAlaPheValIleAlaHisGlnTh		400
1261	CTGAACCCATGAGAGAGGCTTCTCTCAGCGCTTTGGTGGCCATGAGAC		1310
400	rgLysHisValLeuGlyMetGlnHisAspGlyGlnGlyAsnGlyCysAlaA		417
1311	TGGCCACGCTGGGCGCATATGAGACGACGAGGAGGGCAACGCTGTGGCG		1360
417	spGlnThrSerLeuGlySerValMetAlaProLeuValGlnAlaAlaPhe		433
1361	ACGAGAGTGGGCTGGGACACATCATGATGGCGCCCTGTGACAGCCGCTTC		1410
434	HisArgPheHisTrpSerArgCysSerLysLeuGlnLeuSerArgTyrLe		450
1411	CAACGGTTCCTACTGGTCCCGCTGCAGACAGCAGGATGAGGCGCTTACT		1460
450	upProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTrpProG		467
1461	GCACCTCTTATGACTGCTCTGTGATGATACCCCTTGGCCACGACTGGCCG		1510
467	lnProProGluLeuProGlyIleAsnTyrSerMetAlaPheGlnCysArg		483
1511	CGCTGCCACAGCTCCCGGAGACATGCACTACTCATGACAGCAAGCAATGCCG		1560
484	PheAspPheGlySerGlyTyrGlnThrCysLysLeuAlaPheArgThrPheG		500
1561	TTTGACTCTGGCGCTGGGCTACATGATGATGACAGGGGTTCGCGACTTTTGA		1610
500	upProCysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCysL		517
1611	CCCTCTCAAGCAGCTGTGGGACGACATCTTACAAACCCCTACTTTTGA		1660
517	ysThrLysLysGlyProProLeuAspGlyThrGlnCysAlaProGlyLys		533
1661	AGACCAAGAAAGGGGCCCTTGGAGAGGAGATATGCTGACACTGGCAAG		1710
534	trpCysPheLysGlnHisCysIleTrpLysSerProGlnGlnThrTyrG		550


```

1711 CATGTTTAAAGACACTCATCTGTGGACACCTGAC...ATCCTCA 1757
550 YGlnAspGlyGlyTyrSerSerTrpThrLysPheGlySerCysSerArgS 567
1758 ACGGAGCGAGGAGCGGCGCTTGAGTCCGTTTGGCTCTCTCTCAACGTA 1807
567 eRcysGlyGlyGlyValArgSerArgSerArgSerCysAsnAsnProSer 583
1808 CCGTGGCAGGCGGCGTGAAGTTACAGACCCCGAGTGTACACACCCACAC 1857
584 ProAlaTyrGlyGlyValArgProCysLeuGlyPrometPheGluTyrGlnVa 600
1858 CCGGCGACACCGGGGCGCGACCTCGGCGCTTGCGTACGACTTCCAGCT 1907
600 LcysAsnSerGluGluCysProGlyThrTyrGlnAspPheAlaGlnG 617
1908 CTGACGCGCCGACGAGACTGCGCGACCTCGGCTGCTGCTGCTGCGAGAGC 1957
617 IncysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSer 633
1958 AGTCCGCGCAGTGGAGCCTGTACTTCCAGCAGCGGCGCCGACGACAC 2007
634 TrpValProTyrGluProAspAspAlaGlnLysCysGluLeuLeuLeuCy 650
2008 TGCCTGCGCCCGACGAGCAGCCGCGATGCCAAGAGATGCCACCTGACTG 2057
650 sGlnSerAlaAspThrGlyAspValValPheMetAsnGlnValValHisA 667
2058 CAGGTCAGGAGGAGACCGGGAGGTGTGTCCATGAAGCGCATGTGGTCA 2107
667 spGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArgGly 683
2108 ATGGGACGGCGTGTCTCTACAGAGAGCGCTTCCAGCTGTGTGTGCGCGG 2157
684 GluCysValProValGlyLysAspLysGluValGlySerMetLysAlaAs 700
2158 GACTGTGAGAGAGGTGGCGTGTACGCTGTGCTGCTGCTGACGAGCAGGA 2207
700 PAspLysCysGlyValCysGlyGlyAspAsnSerHisCysArgThrAla 717
2208 ACACAAGTGTGGCGTGTGCGGAGGAGACACACCGCAGCTGCAAAATGGTGA 2257
717 yGlyThrLeuGlyLysAlaSerLysGlnAlaGlnValLeuLysLeuVal 733
2258 AGGGACGCTTCACACGCTACCCCAAGCATGCTGTACATCAAGATGTTT 2307
734 GlnIleProAlaGlyAlaArgHisIleGlnIleGlnAlaLeuGlnLysSe 750
2308 GAGATCCCTGCAGAGGACGACACCTGCTCATTCAGAGAGTACAGCGCCAC 2357
750 TrpHisArgSerValValLysAsnGlnValThrGlySerPheIleLeuA 767
2358 CAGCCACCATCTCGCGCTCAGAGACCTGAGACAGCGCAAGTTCATCTTA 2407
767 snProLysGlyLys...GlnAlaThrSerArgTrpPheMetAlaMetGly 782
2408 ATGAAGAAATGACGTGATGTCAGTTCCAAACCTTCATTTGCATGGCGG 2457
783 LeuGluTrpGluAspAlaValGluAspAlaLysGlnSerLeuLysThrSe 799
2458 GTGAGTGTGGAGTACAGAGAGAGAGCGCGGAGAGCGTGCAGACCAT 2507
799 rGlyProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGluG 816
2508 GGGGCCCTCCACGCGACATACACCTTCTGTGTCATCCG.....GTGG 2551
816 LysGlyProArgSerSerLeuAlaTyrLysTyrValIleHisGluAspLeu 832
2552 GAGACACCGCGGTCTACGTACGTACAAATACATGATCTCATGGGAGCTCA 2601
833 LeuProLeuIleGlySerAsnAsnValLeuLeuGluGluMetAspThrTy 849
1711 CATGTTTAAAGACACTCATCTGTGGACACCTGAC...ATCCTCA 1757
550 YGlnAspGlyGlyTyrSerSerTrpThrLysPheGlySerCysSerArgS 567
1758 ACGGAGCGAGGAGCGGCGCTTGAGTCCGTTTGGCTCTCTCTCAACGTA 1807
567 eRcysGlyGlyGlyValArgSerArgSerArgSerCysAsnAsnProSer 583
1808 CCGTGGCAGGCGGCGTGAAGTTACAGACCCCGAGTGTACACACCCACAC 1857
584 ProAlaTyrGlyGlyValArgProCysLeuGlyPrometPheGluTyrGlnVa 600
1858 CCGGCGACACCGGGGCGCGACCTCGGCGCTTGCGTACGACTTCCAGCT 1907
600 LcysAsnSerGluGluCysProGlyThrTyrGlnAspPheAlaGlnG 617
1908 CTGACGCGCCGACGAGACTGCGCGACCTCGGCTGCTGCTGCTGCGAGAGC 1957
617 IncysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSer 633
1958 AGTCCGCGCAGTGGAGCCTGTACTTCCAGCAGCGGCGCCGACGACAC 2007
634 TrpValProTyrGluProAspAspAlaGlnLysCysGluLeuLeuLeuCy 650
2008 TGCCTGCGCCCGACGAGCAGCCGCGATGCCAAGAGATGCCACCTGACTG 2057
650 sGlnSerAlaAspThrGlyAspValValPheMetAsnGlnValValHisA 667
2058 CAGGTCAGGAGGAGACCGGGAGGTGTGTCCATGAAGCGCATGTGGTCA 2107
667 spGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArgGly 683
2108 ATGGGACGGCGTGTCTCTACAGAGAGCGCTTCCAGCTGTGTGTGCGCGG 2157
684 GluCysValProValGlyLysAspLysGluValGlySerMetLysAlaAs 700
2158 GACTGTGAGAGAGGTGGCGTGTACGCTGTGCTGCTGCTGACGAGCAGGA 2207
700 PAspLysCysGlyValCysGlyGlyAspAsnSerHisCysArgThrAla 717
2208 ACACAAGTGTGGCGTGTGCGGAGGAGACACACCGCAGCTGCAAAATGGTGA 2257
717 yGlyThrLeuGlyLysAlaSerLysGlnAlaGlnValLeuLysLeuVal 733
2258 AGGGACGCTTCACACGCTACCCCAAGCATGCTGTACATCAAGATGTTT 2307
734 GlnIleProAlaGlyAlaArgHisIleGlnIleGlnAlaLeuGlnLysSe 750
2308 GAGATCCCTGCAGAGGACGACACCTGCTCATTCAGAGAGTACAGCGCCAC 2357
750 TrpHisArgSerValValLysAsnGlnValThrGlySerPheIleLeuA 767
2358 CAGCCACCATCTCGCGCTCAGAGACCTGAGACAGCGCAAGTTCATCTTA 2407
767 snProLysGlyLys...GlnAlaThrSerArgTrpPheMetAlaMetGly 782
2408 ATGAAGAAATGACGTGATGTCAGTTCCAAACCTTCATTTGCATGGCGG 2457
783 LeuGluTrpGluAspAlaValGluAspAlaLysGlnSerLeuLysThrSe 799
2458 GTGAGTGTGGAGTACAGAGAGAGAGCGCGGAGAGCGTGCAGACCAT 2507
799 rGlyProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGluG 816
2508 GGGGCCCTCCACGCGACATACACCTTCTGTGTCATCCG.....GTGG 2551
816 LysGlyProArgSerSerLeuAlaTyrLysTyrValIleHisGluAspLeu 832
2552 GAGACACCGCGGTCTACGTACGTACAAATACATGATCTCATGGGAGCTCA 2601
833 LeuProLeuIleGlySerAsnAsnValLeuLeuGluGluMetAspThrTy 849
2602 CTGAAT...GTGATGACAAACAGTCTGGAAGAGAGACTGTGTGATGTA 2648
849 rGluTrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGlyG 866
2649 CAGTGGGCGCCCTCACAAGAGTGTCTCGTCTCCAGACCCCTGTGGCGAG 2698
866 LylIleGlnPheThrLysTyrGlyCysArgArgArgArgAspHisMet 882
2699 GTTCCAGCTTCCACCAAGTATGCTGCGCGGAGAGGTGGAGACCAAGATG 2748
883 ValGlnArgHisLeuCysAspHisLysLysArgProLysProIleArgAr 899
2749 GTACACGGTGGCTTCTGTGCGCGCCCTCTCGAAGCCCAAGCATCCGAG 2798
899 gArGcysAsnGlnHisProCysSerLlnProValTrpValThrGluTr 916
2799 AGCGTCAACACCAACAGAGATGCTCCAGCAGCACTGTGGGTACAGCGCAT 2848
916 TrpGlyLacysSerArgSerCysGlyLysLeuGlyValGlnThrArgGly 932
2849 GGGAGCCATGTAGCCAGACCTGTGGCGGAGACGACATCAGAGTGGCTCC 2898
933 lIleGlnCysLeuLeuProLeuSerAsnGlyThrHisLysValIleProAl 949
2899 GTGCGCTGCATTACCGCGCTTACAGACACACACACCCCTCGCGACACG 2948
949 aLysAlaCysAlaGlyAspArgProGluAlaArgArgProCysLeuArgV 966
2949 CAAGCACTCATGTAGACCGCGCGCGAGACCGCGCGCGCGCGCGCGCGG 2998
966 aLProCysProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAla 982
2999 ACCTGCGCCCTGTGTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGAG 3048
983 ThrCysGlyGluGlyTlIleGlnGlnArgGlnValValCysArgTrpAsnAl 999
3049 ACCTGTGGGAGAGCGACCCAGAGCGCGCACTGCTGCGCACCGCGCGGA 3098
999 aAsnSerLeuGlnHisCysGlyLysAspArgProAspThrValGlnValC 1016
3099 CGACAGCTTCGCGCATCTGCCAGAGAGCGCTGAGACAGCGAGGAGCCT 3148
1016 ySerLeuProAlaCysGlyLysGlnHisGlnAsnSerTrpAlaArgAla 1032
3149 CGAGCTTGCGCGCGCTGCTCCGGAACATCTCAGATCCCTCCACAGAAAGC 3198
1033 AspValTrpGluLeuGlyThrProGluGlyGlnTrpValProGlnSerG 1049
3199 TACGTA.....GTTCAGTGGCTGTGCCCGCGCGGA 3227
1049 yProLeuHisProIleAsnLysIleSerSerMetCysAlaAlaGluProC 1066
3228 CCGCGACTGCGCATCCGGAAGATCTGTCA.....AAGCGCCACT 3268
1066 yThrGlyAspArgSerValPheCysGlnMetGlnValIleuAspArgTyr 1082
3269 GCCAAGGCGCAATCAATATCTGTAGATGGAAGTCTTGTGCCCTAT 3318
1083 CysSerIleProGlyTyrHisArgLeuCysCysValSerCys..... 1096
3319 TGTTCATCTCCAGGCTCAACAAGCTGTGCTGCAAGTCTGTAACTGTA 3368
1097 .....lIleLysLysAlaSerGlyProAsnProGlyPro. 1107
3369 CAACAACCTCAGCAGCTGAGGCGAGATAGAGCCACCGCTGGGAGAC 3418
1108 .....AspProGlyProThrSerLeuProProPheSerThrPro 1120
3419 ACAACGACATTTAGAGTGTATGCTACCTCCACAGTGGCCACTGTAGGC 3468
1121 GlySerProLeuProGlyProGlnAspProAlaAspAlaAlaGluProB 1137
3469 ATGAGGTGGCGCATCACCAGACCCCTGGAGGTGCTCTCAATGC 3518

```


264 LeuLeuValAlaAspSerValValArgPheHisGlySerGluHisVal 280
 898 CTGGCGGCGCTGGATGACTCTGTGTGCAAGTCCACGGGAGGACGCGT 947
 280 IGlAsnTyrValLeuThrLeuMetAsnIleValAspGluIleTyrHisA 297
 948 ACAGAAAGTACTGCTGACACTCATGAACTTGTCAATGAATCATCATG 997
 297 SPGLuSerLeuGlyValHisIleAsnIleAlaLeuValArgLeuIleMet 313
 998 ACGAGTCTTGGTGGCCACATCAACGTGGTCCGTGGTGGATCATCTTC 1047
 314 ValGlyTyrArgGlnSerLeuSerLeuIleGluArgGlyAsnProSerAr 330
 1048 CTGAGCTATGAAAGCTCCATGAGCTTATGAGATGGGAGACCTCTTCA 1097
 330 GSerLeuGluGlnValCysArgTTPAlaHisSerGlnGlnArgGlnAsp 347
 1098 GAGCTGGAGAAATGTCTGCCGTGGCTACCTCCAGCAACCCAGACA 1147
 347 rSerHisAlaGlnHisAspHisValValPheLeuThrArgGlnAsp 363
 1148 CGGGCCACGATGAATACACGATCACGCGCATCTCTCACACGCGAGAC 1197
 364 PheGlyProSer.....GlyTyrAlaProValThrGlyMetCysH 377
 1198 TTTGGGCTTCCGGCATGCAAGGCTATGCCCTGCTACCGCGCATGTGGCA 1247
 377 sProLeuArgSerCysAlaLeuAsnHisGluAspGlyPheSerSerAlaP 394
 1248 TCCGCTCCGAGCTGCACCTGAAACCATGAGACGGCTTCTCTCAGCCT 1297
 394 heValIleAlaHisGlyThrGlyHisValLeuGlyMetGluHisAspGly 410
 1298 TTGTGGTGGCCATGAGACTGGCCACGTGTGGCATGAGACGACGCGG 1347
 411 GlnGlyAsnGlyCysAlaAspGluThrSerLeuGlySerValMetAlaP 427
 1348 CAGGGCCACCGCTGTGGCGAGGTGGCGGTGGCGAGCATCATGGCGCG 1397
 427 oleuValGlnAlaAlaPheHisArgPheHisIleTyrPheArgCysSerLysL 444
 1398 CTTGGTGGACGGCGCTTCCACCGCTTCCACGTGGTCCCGCTGACACGAC 1447
 444 euGluLeuSerArgTyrLeuProSerTyrAspCysLeuLeuAspAspPro 460
 1448 AGGAGCTGAGCGCTACCTGCACTGCTATGACTGCTGTGATGATACCC 1497
 461 PheAspProAlaThrProGlnProProGluLeuProGlyIleAsnTyrSe 477
 1498 TTGGCCACGACTGGCGCGCTGGCCAGCTCCGCGGACTGCACTACTGC 1547
 477 rMetAspGlnGlnCysArgPheAspPheGlySerGlyTyrGlnTyrCysL 494
 1548 CATGAACGAGCAATGGCGCTTGTGACTGTGGCTGGCTACATGATGTCCA 1597
 494 euAlaPheArgThrPheGluProCysLysGlnLeuTyrCysSerHisPro 510
 1598 CGCGCTTCCGAGCTTTGACCCCTGCAAGCAGCTGTGGTGCAGCATCT 1647
 511 AspAspProTyrPheCysLysThrLysLysGlyProProLeuAspGlyTh 527
 1648 GACAAACCTTACTTTTGGCAAGACCAAGAGGGGCCCTTGGAGCGGAC 1697
 527 rGluCysAlaProGlyLysTyrCysPheLysGlyHisCysIleTyrLys 544
 1698 TATGTGTGCACTGGCAAGCATTTGTTAAAGACATGATGTGGCGCA 1747
 544 exProGluInThrTyrGlnAspGlyGlyTyrPheSerTyrThrLys 560
 1748 CACCTGAC...ATCCTCAACGCGAGCGAGCTGGGCGCTTGGAGTCCG 1794

561 PheGlySerCysSerArgSerCysGlyGlyValAlaArgSerArgSerAr 577
 1795 TTGGCTCTCTCTCAGCACTGTGGACAGGCGGTGAATTCAGAACCCG 1844
 577 GSerCysAsnAsnProSerProAlaTyrGlyArgProCysLeuGlyP 594
 1845 CCAGTGTACAAACCCACACCGCGGCAACGGGGGCGCGCACTGCTCGGCGC 1894
 594 rMetPheGluTyrGlnValCysAsnSerGluGluCysProGlyThrTyr 610
 1895 TTGGCTACGACTTCTGACCTGTGACGCGCCAGACATCTCCGACTCTCG 1944
 611 GluAspPheArgAlaGlnGlnCysAlaLysArgAsnSerTyrTyrAlaH 627
 1945 GCTACTTCCGCGAGGAGAGAGTCCGCGGCACTGGGACCTGACTTTCAGCA 1994
 627 sGlnAsnAlaLysHisSerTyrValProTyrGluProAspAspAlaG 644
 1995 CGGGGAGCGCCAGCACCTGCTGCCCGCCACGAGCACCGGATGCCAAG 2044
 644 LnyCysGluLeuIleCysGlnSerAlaAspThrGlyAspValPhe 660
 2045 AGAGATGCCAAGCTGTACTGTGAGTCCAGAGGACCGGGAGGTGTCTCC 2094
 661 MetAsnGlnValHisAspGlyThrArgCysSerTyrArgAspProTyr 677
 2095 ATGAAGGCATGTGTGATGATGGAGCGCTGCTCTACAAAGACGCTT 2144
 677 rSerValCysAlaArgGlyGluCysValProValGlyCysAspLysGluV 694
 2145 CAGCTCTGTGTGGCGGAGCTGCGAGAAAGGGGCTGTGACGCTGTGA 2194
 694 alGlySerMetLysAlaAspAspLysCysGlyValCysGlyLysAspAsn 710
 2195 TCGGCTTCCACAGCAGCAAGAACATGTTGGCGTGGCGGAGGGAGCAAC 2244
 711 SerHisCysArgThrValLysGlyThrLeuGlyLysAlaSerLysGlnAl 727
 2245 AGCGACTGCAAAAGTGTCAAGGGCACGTTTCACACGCTCACCCAAACCA 2294
 727 acGlyAlaLeuLysLeuValGlnIleProAlaGlyAlaArgHisIleGln 744
 2295 TGGTTACATCAAGATTTTGAATGATCCCTGACAGACCCAGACCTGCTCA 2344
 744 IeGluAlaLeuGluLysSerProHisArgSerValValLysAsnGlnVal 760
 2345 TTCAGGAGGTAGACGCCACCCACCATCTGTGGCGCTCAAGAACCTGGAG 2394
 761 ThrGlySerPheIleuAsnProLysGlyLys...GluAlaThrSerAr 776
 2395 ACAGGCAAGTTCATCTTAAATGAAGAAATGACGTGATGCCAGTTCCAA 2444
 776 gThrPheThrAlaMetGlyLeuGluTTPGluAspAlaValGluAspAla 793
 2445 AACCTTCATTTGCCATGGCGGTGAGTGGAGTACAGACAGCGAGCGGCC 2494
 793 ySerLeuLeuLysThrSerGlyProLeuProGluAlaIleAlaIleLeu 809
 2495 GGGAGACGCTGACAGACCATGGCGCCCTTCCACGGGACCATACCGTTCTG 2544
 810 AlaLeuProProThrGluGlyLysProArgSerSerLeuAlaTyrLys 826
 2545 GTCATCCGCG...GTGGAGACACCGCGGTCTACCTGACGTGCAATA 2588
 826 rValIleHisGluAspLeuProLeuIleGlySerAsnValLeu 843
 2589 CATGATTCATGAGACTCATCTGAAT...GTGATGACAAACATCTCTCG 2635
 843 euGluGluMetAspThrTyrGluThrPalaLeuLysSerTTPAlaProCys 859
 2636 AAGAGACTCTGTGTCTACAGAGTGGCGCTGAAAGAGTGTCTCCGTGC 2685
 860 SerLysAlaCysGlyGlyIleGlnPheThrLysTyrGlyCysArgAr 876

Thu Aug 29 10:45:52 2002

us-09-938-330-20.rng

Page 27



OM of: US-09-938-330-20 to: EST:★ out_format : pfs
Date: Aug 28, 2002 9:01 PM

About: Results were produced by the Gencore software, version 4.5
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=lrframe-p2n.model -DEV=xlP
-Q/cgu2.1/USPTO.spool/US0938330/r/unat.26082002.15364.23130/app_query.fast.1.13330
-BD=EST -OEM=fastap -SURF=fast -GAOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -QGAOP=4.500
-QGAPEXT=0.050 -XGAOP=10.000 -XGAPEXT=0.500 -FGAOP=6.000
-FGAPEXT=7.000 -YGAOP=1.000 -YGAPEXT=0.500 -DELCP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=humand0.cdi
-ST=45 -DOCALL=200 -THR_SCORE=pct -THR_MAX=10 -THR_MIN=0
-IGN=10 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEADS=SIZE=500
-NLEN=0 -MAXLEN=2000000000 -USER=US0938330.@CGN1_1.6777
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLIPXY -WAIT -THREADS=1
```

Search information block:

```
Query: 05-09-938-330-20
Query length: 1252
```

Database sequences: 13736207
Database length: 1841457050

Search time (sec): 2799.900000

score_list:

Sequence	Strid	Orig	ZScore	Score	Len	Documentation
gb_est2:BF123774	+	1078.00	1156.20	3.7e-55	863	! BF123774 6017600281 NCI CGAP-
gb_hlc:BF003369	+	933.50	1010.12	5.1e-47	2931	! BC003269 Mus musculus, Similia
gb_est2:BF906533	+	862.50	928.33	1.8e-42	774	! BF906533 t107-070123-111200-501
gb_est1:BF081935	+	860.50	923.92	3.2e-42	724	! BF081935 602877161F1 NCI CGAP
gb_est1:BF655886	+	824.50	897.46	3.4e-40	564	! BF655886 BB635886 RIKEN full-1.1
gb_est2:BF381511	+	808.50	870.48	3.1e-39	562	! BF381511 6012725204F1 NCI CGAP
gb_est2:BF1858020	+	797.50	855.56	2.1e-38	757	! BF1858020 6033848425F1 NIH_MGC_8
gb_est2:BF1930891	+	769.50	829.11	6.2e-37	529	! BF1930891 237553 MARC 2Pig Sus
gb_est1:BF043134	+	767.00	828.76	6.5e-37	418	! BF043134 A245533 MARC 2Pig Sus
gb_est2:BF096904	+	748.00	805.51	1.3e-35	555	! BF096904 B1096504 NIBB Mochi
gb_hlc:BF013707	+	746.00	798.81	1.3e-35	878	! BC013707 Homo sapiens, clone J1
gb_est2:BF1350958	+	731.50	768.09	1.2e-34	481	! BF1350958 237691 MARC 2Pig Sus
gb_est1:BF906535	+	710.00	765.89	2.1e-33	544	! BF906535 t107-070123-111200-501
gb_est1:AL046910	+	699.00	755.15	4.1e-33	427	! AL046910 DKFZP868081F1 rat 566
gb_est1:AW657341	+	688.50	741.65	4.8e-33	546	! AW657341 109954 MARC 1BOV Bos
gb_est1:AL0688776	+	678.50	721.66	2.8e-32	756	! AA920219 v454909.4.1 Stratagene
gb_est1:AL0688776	+	673.50	723.70	4.6e-31	655	! AL0688776 AL648772 XGC-gastrula
gb_est2:BF026102	+	628.00	674.49	2.5e-28	671	! BF026102 6033847256F1 NIH_MGC_8
gb_est2:BF1860076	+	617.50	665.27	8.3e-28	545	! BF1860076 60285973F1 NIH_MGC_1
gb_est2:BF1409671	+	600.50	645.26	1.6e-26	877	! BF1409671 60296265F1 NCI CGAP
gb_gss:CN054198	+	593.00	639.13	2.4e-26	533	! BF175928 46856 MARC 2BOV Bos
gb_est2:BF057512	+	587.00	632.93	7.7e-26	974	! AL270720 Tetrahodon nigroviridis
gb_est2:BF1092940	+	585.50	628.00	8.4e-26	755	! BG575152 60259827F1 NIH_MGC_8
gb_est2:BF906528	+	580.50	625.05	1.4e-26	726	! BF906528 60285973F1 NIH_MGC_1
gb_est2:BF14345878	+	577.00	620.94	2.4e-25	568	! BF14345878 60285973F1 NIH_MGC_1
gb_est2:BF14345883	+	577.00	620.92	2.4e-25	588	! BF14345878 374838 MARC 2Pig Sus
gb_est2:BF8382289	+	570.00	618.61	3.4e-25	589	! BF8382289 374838 MARC 2Pig Sus
gb_est2:BF0588231	+	566.00	610.85	8.9e-25	493	! BF8382289 374838 MARC 2Pig Sus
gb_est2:BF1339744	+	552.50	596.33	5.7e-24	493	! BG3828231 298056 MARC 1Pig Sus
gb_est1:BF128052	+	534.50	575.84	7.9e-23	552	! BF1339744 364841 MARC 2Pig Sus
gb_est1:AA832579	+	528.00	573.28	1.1e-22	353	! BF1339744 364841 MARC 2Pig Sus
gb_est2:BF753141	+	503.00	542.48	5.7e-21	523	! AA832579 DEPA1801 Rat Lambda 2Pig
gb_est1:AV614541	+	500.00	540.48	7.4e-21	488	! AA832579 DEPA1801 Rat Lambda 2Pig
gb_est1:AV666539	+	496.50	534.95	9.7e-21	575	! BF753141 206054 MARC 2BOV Bos
gb_est2:BF537719	+	468.50	499.72	1.4e-18	924	! AV614541 AV614541 Bos taurs ad
gb_est2:BF137395	+	457.00	489.47	5.1e-18	746	! AV666539 AV666539 Bos taurs ad
gb_est2:BF1984736	+	445.50	487.26	6.8e-18	460	! BF537719 AV666539F1 NIH_MGC_10
gb_est2:BF0940741	+	448.50	483.15	1.2e-17	561	! BF1984736 603331940F1 NIH_MGC_94
gb_est2:BF0940741	+	448.50	483.15	1.2e-17	561	! BF0940741 Fui0410. y3 Campbelli
gb_est2:BF0940741	+	448.50	483.15	1.2e-17	561	! BF0940741 Fui0410. y3 Campbelli
gb_est2:BF0940741	+	448.50	483.15	1.2e-17	561	! BF0940741 Fui0410. y3 Campbelli
gb_est2:BF0940741	+	448.50	483.15			

9b_est2:BI153886	+	445.00	478.17	2.2e-17	634	!	BI153886	602870889FL	NC1	CGAN
9b_est1:AM483252	+	444.00	481.01	1.5e-17	427	!	AM483252	51399	MARC	2PG
9b_est2:BF033017	+	443.00	472.05	4.8e-17	945	!	BF033017	601455701FL	NH	MCCG
9b_est2:BI142865	+	431.00	467.55	8.5e-17	405	!	BI142865	3708299	MARC	2PG
9b_est2:BI103177	+	422.00	450.52	7.6e-16	849	!	BI103177	602889417FL	NC1	CGAN

seq_name: gb_est2:BF123774

```
seq_documentation_block:
  LOCUS      PF133774
```

DEFINITION 601760028F1
mRNA sequence

ACCESSION	BF123774
VERSION	BF123774.1

NEIMOWUDS	EST.
SOURCE	house mouse.

was included
Eukaryota; M

REFERENCE 1 (bases 1

TITLE	National Ins

COMMENT Contact: Rob

Tissue Procurement

CDNA Library

Clone distr
found + brown

http://image
Plate: T1AM9

	high quality
FEATURES	Loc

or

 $\text{d}b/\text{d}p$

/c1

/de

/ no

Lib.

BASE COUNT	167 a	NLH
------------	-------	-----

ORIGIN

alignment_scores:

Ratio: 4.3

US-09-938-330-20 x BF123

Align seg 1/1 to: BF123

```
438 TrpSerArgCysSerL
```

2 1GG1C1CGC1GGAGCCT

.....

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

```

102 TGGCCGGGATGACATCTCCATGATGACCAATGCCGGTTGACCTTGGC 151
488 SerGlyThrGlnThrCysLeuAlaPheArgThrPheGluProCysLysG1 504
152 ACTGGGTACACACCTGCTTACCTTACGACCTTGGACCTTGGACAGCA 201
504 nleuTrpCysSerHisProAspAsnProTyrPheCysLysThrLysLysG 521
202 GCTATGATGGAGCCACCCCTGACCAACCTATCTCTCAAGACCAAGAAAG 251
521 LyrProLeuAspGlyThrGluCysAlaProGlyLysTrpCysPheLys 537
252 GGGCGGGC.CTGGATGGACAGAGTGTGCACAGGCAAGTGTCTTCAAA 300
538 G1HisCysIleTrpLysSerProGluGlnThrTyrGlyGlnAspGlyG1 554
301 GGCCA.TGATCTGGAAAGTCACAGACCAAACTTATGCCAGATGAGAG 349
554 YTrpSerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyG 571
350 CTGGAGTTCCTGGACCA..TTTGGCTCATGTCTCGTGTGGAGAGAG 398
571 LysAlaGSerArgSerArgSerCysAsnAsnProSerProAlaTyrGly 587
399 GGGTGCATCCCGAAGCCGAGCTGGACAACTCTCTCCAGCTATGGA 448
588 G1YArgProCysLeuGlyProMetPheGluTyrGlnValLysAsnSerG1 604
449 GGCGCGCCGTGCTCAGGCTGATGTTGAGTACCAAGATGCAACACTGA 498
604 uGluCysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysA 621
499 GGAAGTCTCTGGGCTTACAGAGACTTCCAGCCCACTGTGCCAAGC 548
621 rGAsnSerTyrTyrValHisGlnAsnAlaLysHisSerTrpValProTyr 637
549 GAACTCTACTATATCCACAGATGCACAAAGCAG.TGGTGGCTATG 597
638 GluProAspAspAlaGlnLysCysGlnLeuIleCysGlnSerAlaAs 654
598 AGCCCGACAGGTGATGCCACAGATGCGAGCTCATTTCA...GTCTCGA 644
654 PThrGlyAspValAlaPheMetAsnGlnValAlaHisAspGlyThrArgC 671
645 CACTGGCATATGCTCTCATGAACAGTTGCC.CAACAGGAGACAGCT 693
671 ySerTyrArgAspProTyrSerValCysAlaArgGlyGluCysValPro 687
694 GAGTATCGCGAAC...TTAAGGTGTGGGCC...GGGAGTGTTCCT 737
688 ValGlyCysAspLysGluValGlySerMetLysAlaAspLysCysG1 704
738 .....GG 739
704 YValCysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeu. 720
740 TGGCGCGGGGGGGCC.....CGCACGCTGCAGAGGGGGGCTCC 777
721 Gly.....LysAlaSerLysGlnAlaGlyAlaLeuLysLeuVa 733
778 GGGGACACCCCGCACCGGGGGGGGGGGGGGGTCT..... 816
733 LcInIleProAlaGlyAlaArgHisIleGlnIleGlnAlaLeuGluLys 750
817 .....CGGGCCGGGGCGGGCCCAACAGGA..... 840
750 erProHis 752
841 ..CCCCAC 846
seq_name: gb_hnc:BC003269
seq_documentation_block:

```

```

LOCUS      BC003269                2931 bp      mRNA           linear      HTC 12-JUL-2001
DEFINITION Mus musculus, similar to a disintegrin-like and metalloprotease
            (repolysin type) with thrombospondin type 1 motif, 1, clone
            IMAGE:349206, mRNA.
ACCESSION  BC003269
VERSION    BC003269.1   GI:14707747
SOURCE     Mus musculus
            house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 2931)
AUTHORS   Strausberg, R.
TITLE     Direct Submission
JOURNAL    Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK     NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT    Contact: MGC help desk
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalona@bcm.tmc.edu,
            Villalón, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia,
            A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
            Muzny, D.M., Gibbs, R.A.
FEATURES   source
            1..2931
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:349206"
            /tissue_type="Mammary tumor, C3(1)-tag model. Infiltrating
            ductal carcinoma. 5 month old virgin mouse."
            /clone_id="NCI_CGAP_Mam6"
            /lab_host="DH10B"
            /note="Vector: pCMV-Sport6"
BASE COUNT      708 a      812 c      811 g      600 t
ORIGIN
alignment_scores:
    Quality: 953.50      Length: 913
    Ratio: 1.884      Gaps: 35
    Percent Similarity: 55.422      Percent Identity: 29.025
alignment_block:
US-09-938-330-20 x BC003269 ..
Align seg 1/1 to: BC003269 from: 1 to: 2931
3 ProleuArgAlaLeuLeuSerTyrLeuLeuProLeuHisCysAlaLeuGly 19
111 :::: 111 :::: 111 :::: 111 ::::
322 CCAGCGAAGCTTACGCGTCTAAACGTATATCCAGACCAAGAGTTCTTG 371
19 sthAla...AlaGlySerArgThrProGluLeuHisLysSerGlyLys. 34
111 :::: 111 :::: 111 :::: 111 ::::
372 CGGTGCGAGGTGGGTGCGATGCAAGCAAGATCCCTTGGGTGACGCA 421
35 .....LeuSerAspTyrGlyValThrValProCysSerThrAsp 47
111 :::: 111 :::: 111 ::::

```



```

422 AGCAGAACCCCTGCTCCGACATGGGG...GACGTCCAGCGGGCAGCAGAGA 468
48 PheargglyIyArpheLeuSerHisValValSerGlyProAlaAlaAla 64
469 TCTGGGGCTCTCTGTCCGACACATGCTGTG..... 501
64 rAlaGlySerMetValValAspThrProThrLeuProArgHisSers 81
502 .....CTGCTCTGCTTCCATTAACATGCTGCTATGTGGCGGGGGG 544
81 erHisLeuArg.....Val 85
545 CACACAGCGGGCCCCCAGAGAAAGATGAGAGCTGCTGCTGCCCTGG 594
86 AlaArgSerProLeuHisProGlyIyThrLeuTProGlyIyArgValG 102
595 GACGGGGCCCCGGCCAGATTCACACC..... 624
102 yArGHisSerLeuTyrrPheAsnValThrValPheGlyIySLeuHisL 119
625 .....ACACGCTTCGTCTGGAGCCCTTGGCCACGACTACATC 664
119 euArGLeuArgProAsnArgArgLeuValValProGlySerSerValG 135
665 TGAAGTTTGACCGCGACAGCGGCTTCTGGCGCTTGCCCTCACCC 714
136 TrpGlnGluAspPheArgGluLeuPheArgGlnProLeuArgGln.... 150
715 .....ACTGTGGGGCGACGCTCCGGGTCGAGGACCA 746
151 .....GluCysValTyrrThrGlyIy 158
747 ACATGTGACCCCGGAGGACCTGCTACGCTTCTACTCTGGACGG 796
158 aThrGlyMetProGlyIyAlaValAlaValIleSerAsnCysAspGlyLeu 174
797 TGAAGGTCATCCCGCTCTGCGGACGCTCAGCTCTGTGAAGGTGTG 846
175 AlaGlyLeuIleArgThrAspSerThrAspPheIleGlnPro.... 189
847 CCTGTGCTTCTTACCTACAGAGAGAGATTCTTCAATCGACAGCGCC 896
190 .....LeuGlnArgGlyGlnGlnGlnGlnGlnGlnGlnGln 200
897 TGGAGTGGCCACCGAGCGCTGCCCTGCGCTGCCGAGGAGAGATCAT 946
200 erGlyArgThr.....HisValValTyrrArgArgGln..... 210
947 CCGCAGCGGCCGACATGTCACATCTGAGAGCGAAGCGCGGCGCATG 996
211 .....AlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 223
997 GCGCGCAAGTGGCGCTCATGAGCAGAGACCTGCCAACACAGCGACT 1046
223 uHisAsnGlnAla..... 227
1047 GCGACCCGAGAGCCAGAACCCGGAACCACTGCTGCCGGAGCCCA 1096
228 .....PheGlyLeuGlyAspLeuProAsnLeu 236
1097 CCGCTCAGGAGCGGGAAGCCATCAGGACCAAGAACATA..... 1137
237 LeuGlyLeuValGlyAspGlnLeuGlyAspThrGlnArgArgArg 253
1138 .....AGGAAGAGGAGATT 1151
253 sAlaIySProGlySerTyrrIleGlnValLeuLeuValValAspAs 270
1152 TGTGTCCAGCCCCGTTAT...GTGGAACCATGCTGTGGCTGACCAT 1198
270 erValValArgPheHisGlyIySLeuHisValGlnAsnTyrrValLeu 286
1199 CCATGGCGGACTTCCAGCG...AGCGTCTAAAGCATTAACCTTTAAC 1245
287 LeuMetAsnIleValAspGlnIleTyrrHisAspGluSerLeuGlyVal 303
1246 CTGTCTCGGTGGCAGCCAGGTTTTCAGACATCCAGCATTAAGAAATTC 1295
303 sIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrrArgGlnSer 320
1296 AATTAGCTGTGTGTGAAGATCTTGCTGATA...TATGAGAGCAGA 1342
320 euSerLeuIleGlnArgGlyAsnProSerArgSerLeuGlnValGly 336
1343 AGGACCCAGAGTACCTCAATGACGCTTCACCTTCGGAATTTCTGC 1392
337 ArgTrpAlaHisSerGlnGlnArgGlnAspProSer.....HisAlaG 351
1393 AACTGG.....CAGAAACACACAAAGCCCCCATGTGACCGGATCCGA 1436
351 uHisHisAspHisValValPheLeuThrArgGlnAspPheGlyProSer 367
1437 GCACATGACACTGCATTCCTGTTCCACAGACGATTTATGTGGCTCCC 1486
368 .....GlyTyrrAlaProValThrGlyMetCysHisPro 378
1487 ACACGTGTACACTCTCGGAGATGCCAGATTTGGACACTGTATGTACCC 1536
379 LeuArgSerCysAlaLeuAsnHisGlnAspGlyPheSerAlaPheVa 395
1537 ACAGAGAGCTGCTCAGTCATAGAGATGATGTTTGCAGGCGGCTTAC 1586
395 lIleAlaHisGlnThrGlnHisValLeuGlyMetGlnHisSpolGln 412
1587 CACAGCCCAAGATTTGGCGCATGTGTTAACATGCGGACAGAT...GATG 1633
412 lYAsnGlyCysAlaAsp.....GluThrSerLeuGlySerValMet 425
1634 CTAAACACTGTCCAGCTTGAATGTGTGACTGGCGATTCATCATGTATG 1683
426 AlaProLeuValGlnAlaAlaPheHisArgPheHisTrpSerArgCys 442
1684 GCTTGATGCTCTCCAGCTTACCATTAAGCAGCCCTGCTCAGCTTCCAG 1733
442 rLysLeuGlnLeuSerArgTyrrLeuProSer.....TyrrAspCysLeu 457
1734 TGCCATACATGTCACAGTCTTCTTATGATTAAGACACAGGGGAATGTTGA 1783
457 euAspAspProPheAspProAlaThrProGlnProProGlnLeuProGly 473
1784 TGGACAAAGCCCAAGATCCATC...AAGTCCCTTGTGATCTTCCCGGT 1830
474 lLeAsnTyrrSerMetAspGlnGlnCysArgPheAspPheGlySerGly 490
1831 ACCTTGTACGATGCCAACCGCAGTGTACATTCATTCGGAAGAGAATC 1880
490 rGlnThrCysLeuAlaPheArgThrPheGlnProCysLysGlnLeuTrp 507
1881 CAAGCATGCGCT.....GATGACCCACACATGACTACCTGTGTGT 1924
507 ySserHisProAspAspProTyrrPhe...CysLysThrLysLysGlyPro 522
1925 GCACGTGCACCTCGGTGGCTTACTGTGTGTCGCAAAACAAACCTTCCCT 1974
523 ProLeuAspGlyThrGlnCysAlaPheGlyLysTrpCysPheLysGln 539
1975 TGGCAGATGTCACACCTGTGGAGAAAGGAATGCTGTCTGAGTGCA 2024
539 sCyslIleTrpLysSerProGlnGlnThrTyrrGlyGln.....AspGlyG 554
2025 GTGGTGAAGAACAGACATGAAGCATTTTCTACTCTGTTATGAGAA 2074
554 lYTrpSerSerTrpThrLysPheGlySerCysSerArgSerCysGly 570
2075 GCTGGGAGCATGGGAGCCTGGGAGACTGTCTCAAGAACCTGTGTGT 2124

```



```

|||||
373 AAATCACCAGAACTCCACGCTGAGGGCCGATCTCGGAACCTTGGGACGC 324
1040 rogluylglntrpyalProglInserglyProleuHisProIleAsnLys 1056
323 CAGAGGGGAGTGGTGGCCACATCTGACCCCTACATCCATTACACAG 274
1057 IleSerMetCysAlaIleagluProCysThrGlyAspAspSerValph 1073
273 ATATCATCA.....ACGAGCCCTGACAGGGAGACAGTCTGCTT 233
1073 eCysglInmetgluValleuAspArgTyrCysSerIleProglYThrHisA 1090
232 CTGCCAGATGGAAGTCTCGATCGCTACTGCTCATTCGCCGCTACACACC 183
1090 rgluCyCysValSerCysTleLysLysAlaSerglyProAsnProgly 1106
182 GCGTCTGCTGTGTCTGTCTGTCATCAAGAGGCGCTGGCCCAACCCCTGGC 133
1107 ProAspProglyProThrSerleuProProPheSer 1118
132 CCAGACTGGGCGCCACCTCACTGCTCTCTCTCG 97
seq_name: gb_estc:BI081935

```

```

seq_documentation_block:
LOCUS BI081935 724 bp mRNA linear EST 20-JUN-2001
DEFINITION 602877161F1 NC1_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5008982 5',
ACCESSION BI081935 mRNA sequence.
VERSION BI081935.1 GI:14500265
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

```

REFERENCE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabrs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L14M11054 row: 1 column: 15
High quality sequence start: 2
High quality sequence stop: 705.
Location/Qualifiers
1. 724

```

```

FEATURES
source
1. 724

```

```

/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5008982"
/clone_id="NC1_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; site_1: SalI;
site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 162 a 208 c 217 g 137 t
ORIGIN

```

```

alignment_scores:
Quality: 860.50 Length: 243
Ratio: 4.413 Gaps: 2

```

```

Percent Similarity: 80.247 Percent Identity: 57.202
alignment_block:
US-09-938-330-20 x BI081935
Align seg 1/1 to: BI081935 from: 1 to: 724

```

```

495 AlapharGthrPhegluProCysLysglInleuThrCysSerHisProAs 511
|||||
1 GCGTCCGAGACCTTGACCCCTGCACAAAGCTCTGGTGACACACCCAGA 50
511 pAsnProIyrPheCysLysThrLysLysGlyProProleuAspLysThrG 528
|||||
51 CAACCCCTACTTTGCAAGACAAAGAGGGTCTCCACTGATGATGACTA 100
528 luCyAlaProglYlYstrPcysPheLysglYHiscysIletrPlySer 544
101 TGTGTGGCCCTGGCAGACAGCTGTTCAAGGCCACATCATCTGACTGACA 150
545 ProgluInThrTyrGlylnAspGlylYtrPserThrPhylyspH 561
|||||
151 CCGCAT..ATCCTCAAAACGAGACGCACTGGGGTCTGCACTCCGTT 197
561 eGlySerCysSerArserCysglYglYlYalArgSerArserArgS 578
|||||
198 TGGCTCCTGCTCACGACGACCTGTGGACAGGCGTGAAGTTCAGACCGCC 247
578 erCysAsnAsnProSerProAlaTyrGlylYalArgProCysLysglYpro 594
|||||
248 AGTGTGACAAACCCACATCCAGCCAAATGGGGCGGCACTGCTGGGCCCTG 297
595 MethGluTyrGlnValCysAsnSerGluGluCysProglYThrTyrGl 611
|||||
298 CCGATGACCTTCACGCTGTCACACCCCAACCCCACTGTCCTGCTGGC 347
611 uAspPheArgAlaGlnInCysAlaLysArgAsnSerTyrTyrValHisG 628
348 TGACTTCGGGAGAGGAGCAGTGTCAAGTGGACCTGTACTTGGACACAG 397
628 lnaAsnAlaLysHisSerTyrPvalProTyrGluProAspAspAlaGln 644
|||||
398 GTGATGTCACAGCACCACTGCTCCTCATATGACACCGGATGCAAGAG 447
645 LysCysglInleuLecysglInserAlaAspThrGlyAspValAlaPheMe 661
|||||
448 AGATGCCACCTTACTGTGTGATGTCACAGAGACGAGGAGCTGTGCTAT 497
661 lAsnGlnValValHisAspGlyThrArgCysSerTyrArgAspProTyrS 678
498 GAAGCGCATGTGTCACAGATGAGACAGCTGTCTCTACAGACAGCCCTCA 547
678 erValCysAlaArgGlyGluCysValProValGlyCysAspLysglVal 694
548 GCGTGTGTGTGCTGGGACCTGACGAGAAAGTGATGTATGCGGTGATC 597
695 GlySerMetLysAlaAspAspLysCysglYalCysglYlYAspAsnSe 711
|||||
598 GGGTCCCGTAAAGCAGCAAGCAAGTGTGGCTGTGTGGGCGCAACAC 647
711 rHisCysArgThrValLysglYThrLeuGlyLysAlaSerIleAla 727
648 CCACTGTACACAGTGTGAAGGCGACATTCACAGGTACACCGCAACAA 697
728 GlyAlaLeuLysleuValGlnIlePro 736
698 GATTACATCACAGATGCTGTGATGCCG 724

```

```

seq_name: gb_estl:BB635886
seq_documentation_block:
LOCUS BB635886 564 bp mRNA linear EST 31-AUG-2001
DEFINITION BB635886 RIKEN full-length enriched, 0 day neonate thymus Mus
musculus cDNA clone A430089F14 5', mRNA sequence.

```

flu late


```

ACCESSION      BE381517
VERSION        BE381517.1 GI:9326882
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
REFERENCE      1 (bases 1 to 562)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/,
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgabs-r@mail.nih.gov
                Tissue Procurement: Gilbert Smith, Ph.D.
                CDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LHAM8787 row: h column: 05
                High quality sequence stop: 562.
FEATURES
    source
        1..562
            location/Qualifiers
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:3601588"
                /clone_lib="NCI CGAP Mam1"
                /tissue_type="tumor, biopsy sample"
                /dev_stage="10 months, virgin"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: PCMV-SHORT6; Site.1: SalI;
                Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT
                library constructed by Life Technologies. Investigator
                providing samples: Gilbert Smith, NIH"
BASE COUNT     106 a 179 c 156 g 121 t
ORIGIN
alignment_scores:
    Quality: 808.50      Length: 187
    Ratio: 4.870         Gaps: 2
    Percent Similarity: 88.770      Percent Identity: 76.471
alignment_block:
    -09-938-330-20 x BE381517 ..
align seg 1/1 to: BE381517 from: 1 to: 562
324 GUAUAGUyAsnProSerArgSerLeuGluGlnValAlaCysArgTrpAlaH 340
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 GAGATTGGAAACCCCTCTCAGAGTCTGGAAGAATGTCTGCCGATGGCCTA 52
340 sserGlnArgGlnAspProSerHisAlaGluHisHisAspHisValY 357
    : ||||| : : ||| : : : : : : : : : : : : : : : : : :
53 CCTCCAGACGAAGCCGGACACAGCACCGATGATGATCCACGACCATGCA 102
357 aPheLeuThrArgGlnAspPheGlyProSer .....GlyTyrAla 370
103 TCTTCTCTCACAGGCGAGATTGTGGCCCTCGGGGCATGCAAGCTATGCT 152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 ProValThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisG 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 CCGGTCACCGGGGATGTGCCACCCCTGTCGGACGCTGATCACTTGAAACATGA 202
387 uAspGlyPheSerSerAlaPheValIleAlaHisGluThrGlyHisValI 404
203 AGATGGCTTCTCTCGGCATTGTGGTGGCTCAACGACGAGCGGCATGTCC 252
404 euGlyMetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSer 420
253 TGGCGATGTGAGCATGACGAGGCAAGGCAACCGCTGCGCTGACAGAGTGCG 302

```

```

421  LeuG1:SerValMetAlaProLeuValGlnAlaAlaPheHisArgPheI 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303  CTGGGACGACATCATGGCCCCCTGGGTACAGAGCGCTTCATCGCTTCCA 352
437  sTrpSerArgCysSerLysLeuGluLeuSerAArgTyrLeuProSerTyrA 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
353  CTGGTCCCGCTGCAGCCAGAGAGAGCTCAAGCCGCTACGTGATTCCTTANG 402
454  spGlysLeuAspAspProPheAspProAlaIrrPro_GlnProProG1 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403  ACGGCGCTGGCGGATGATCCCTTGCCCCACGACGTGGGACACATTGGCTCA 452
470  uLeuProGlyL1LeAsnTyrSerMetAspGluGlnCysArgPheAspPheG 487
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
453  ACGCCGACGGGCTTCATATTCCATGACAGACAGAGTGTGCTTGACTTTGG 502
487  lYserGlyTyrGlnThrCysLeuAlaAlaPheArgThrPheGluProCysLys 503
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
503  GCCTGGGTTACATGATGTGCACGTGCTACCGACCTTCGACCCCTGCAGAA 552
504  GlnLeuTrrp 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
553  CAGCTCTGG 561

seq_name: gp_est2:BI858020

seq_documentation_block:
LOCUS BI858020 767 bp mRNA linear EST 10-OCT-22
DEFINITION 6033848251 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:539351 5',
RNA sequence.
ACCESSION BI858020
VERSION BI858020.1 GI:15998767
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 767)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@ds-remail.nih.gov
Tissue Procurement: DCMD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: L1AM12003 row: a column: 08
High quality sequence stop: 764.
Location/Qualifiers
1..767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:539351"
/clone_lib="NIH_MGC_87"
/lusue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: PCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."
BASE COUNT 239 a 158 c 177 g 193 t
ORIGIN

```

Percent Similarity: 75.000 Percent Identity: 56.818

alignment_block:

US-09-938-330-20 x B1858020 ..

Align seg 1/1 to: B1858020 from: 1 to: 767

```

591 CysLeuGlyPromeTherGluValGlnValCysAsnSerGlnGluCysPr 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    4 TGCCCTGAGTTAAATTTGATGACAGCTTTTAAACACAGGAATGCGCA 53
607 oglyThrTyrluasPheargAlaglnGlnCysAlatysArgAsnSer 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    54 AAACCTTTGAGACTTCAGACACAGCAGTGTGACGACGCAAACTCC 103
624 yTrTyrlValHisGlnAsnAlatysHisSerTrpValProTyrluProAsp 640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    104 ACTTGAATACCAATACCAACACACACTGCTGTCATATGACATCCT 153
641 AspAspAlaGlnLysCysGluLeuLeuLeuGlnSerAlaAspThrGlyAs 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    154 GACCCCAAGAAAGATGCCCTTACTCTCAGTCCAGAGAGAGCTGGAGA 203
657 pValAlaPheMetAsnGlnValValHisAspGlyThrArgCysSerTyra 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    204 TGTGCTTACATGAACAAACACTGTCATGATGAAAGCAGCTGCTTACA 253
674 rGAspProTyrlSerValCysAlaArgGlyGluCysValProValGlyCys 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    254 AATATCATATAGCATATGTGTGCGAGAGAGTGTGTAAGTGGAGCTGT 303
691 AspLysGlnValGlySerMetLysAlaAspAspLysCysGlyValCysGln 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    304 GATAAAGAAATGTGTTCTAATAAGGTGAGATAGTGTGTGTCTGTG 353
707 yGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGlyLysAlas 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    354 AGAGATTAATTCCTCCAGTCCGAAACCTGGAAGGGGACATTTACAGAACTC 403
724 erLysGlnAlaGlyAlaLeuLysLeuValGlnLeuProAlaGlyAlaArg 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    404 CAGGAAGCTGGGTACCTTAAGATGTTGATATACCCCTGGGGCTTAA 453
741 HisIleGlnIleGlnAlaLeuGlnLysSerProHisArgSerValVally 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    454 CATGCTTATATCCAGAGACGAGGCTCTCCTCATATTCCTGCTATTA 503
757 sAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGlnAla 774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    504 GAACCGAGGCTACAGGCCATTATTTAAATGGCAAGGGGAGGAGCA 553
774 hrSerArgThrPheThrAlaMetGlyLeuGlnLysLeuValAlaGln 790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    554 AGTGGGAGCTTTCATAGATCTTGGTGTGAGGGGATATTAACATGAA 603
791 AspAlaLysGlnSerLeuLysThrSerGlyProLeuProGlnAlaIle.. 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    604 GATGACATTTGAAGCTTCCACACCGATGACCTTTACATGACCTGTACT 653
807 .....AlaIleLeuAlaLeuProProThrGlnG 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    654 GCTTGAATTAACCTCAAGAACATGATGAGCGCTCAAGCTGAGACA... 699
816 LysGlyProArgSerSerLeuAlaTyrlLysTyrlValIleHisGlnAspLeu 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    700 .....TATTAATGATCATCTCATGATAAAGACTC 726
833 Leu..ProLeuIleGlySerAsnAsnValLeuLeuGlnGlu 845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    727 TTGTACTTACATCAACACAGCAACATGTATCTCAGAGAGAA 767

```

seq_name: gb_est2:Bf190891

seq documentation_block:

LOCUS Bf190891

529 bp mRNA linear

EST 02-NOV-2000

DEFINITION 237553 MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.

ACCESSION Bf190891

VERSION Bf190891.1

GI:11074260

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 529)

Fahrenkrug, S.C., Frenking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,

Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,

and Keele, J.W.

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 18

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCAGTCACGACG

Plate: 65 row: G column: 12

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

1..529

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 2Pig"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

library made from pooled tissue from testis, ovary,

endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 103 a 163 c 158 g 105 t

ORIGIN

alignment_scores:

Quality: 769.50 Length: 175

Ratio: 4.692 Gaps: 1

Percent Similarity: 93.714 Percent Identity: 78.286

alignment_block:

US-09-938-330-20 x Bf190891 ..

Align seg 1/1 to: Bf190891 from: 1 to: 529

```

270 SerValValArgPheHisGlyLysGlnHisValGlnAsnTyrlValLeuTh 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    3 TCTGTGCTCCAGTCCACAGGAGAGCAGCTGTGAGAACTGCTGCTCAC 52
286 rLeuMetAsnIleValAspGluIleTyrlHisAspGlnSerLeuGlyValH 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    53 TCTCATGACATCGTCAATGAGATTATCATGAGAGCTCCCTGGGGCCC 102
303 isIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrlArgGlnSer 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    103 ACATCAATGATGTCCTGTGGGAGTAATCTGCTGAGATTACGGGAACTCC 152
320 LeuSerLeuIleGlnArgGlyAsnProSerArgSerLeuGlnGlnValCy 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    153 ATGAGCCCATTTGAGATCGGAAACCTCTGACAGTCTGGAGAACTGCTG 202
336 sArgTrpAlaHisSerGlnArgGlnAspProSerHisAlaGlnHis 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

203 CCGCTGGCCCTACCTGCACAGACCGGACACCTGGCCACGAGAGTACC 252
 353 iAspHisValValPheLeuThrArgGlnAspPheGlyProSer..... 367
 253 AGGATCATGTCATTTCTCTCAACCGCGAGACTTTGGCCCTGGGGCATG 302
 368 ...GlyTYrAlaProValThrGlyMetCysHisProLeuArgSerCysAl 383
 303 CAAGGCTATGCTCTCTCACCAGGAGATGTCACACCTGTCGCCACTGCAC 352
 383 aleuasnHisGlnuAspGlyPheSerSerAlaPheValIleAlaHisGluT 400
 353 CTGGAACCAACGAGATGCTCTCTCTCAGCGTTCGTCGTCGCCATGAGA 402
 400 hGlyHisValLeuGlyMetGlnHisAspGlyGlnGlyAsnGlyCysAla 416
 403 CTGGCCATGCTGCTGGCATGAGCATGACGGGACGAGGACCGCTGGGGC 452
 417 AspGluThrSerLeuGlySerValMetAlaProLeuValGlnAlaAlaPhe 433
 453 GACGAGGTCGCGGTCGGCAGCATGATGACACCGCTGCTGCAGGACGCTT 502
 503 CCACGCTTCCACTGCTGCTCCGCTGC 527

seq_name: gb_est1:AJ403134

seq_documentation_block:

LOCUS AJ403134 418 bp mRNA linear (EST 28-JAN-2002)

DEFINITION AJ403134 3.3 (upregulated in larynx carcinoma) Homo sapiens cDNA

ACCESSION AJ403134

VERSION AJ403134.1 GI:9903409

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 418) Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.

AUTHORS Frohme, M., Scharm, B., Delius, H., Knecht, R. and Heideisel, J.D.

TITLE Use of representative difference analysis and cDNA arrays for

JOURNAL transcriptional profiling of tumor tissue

MEDLINE Ann. N. Y. Acad. Sci. 910, 85-104 (2000)

COMMENT Contact: Frohme M

Functional Genome Analysis

Deutsches Krebsforschungszentrum

INF 506, 69120 Heidelberg, Germany

Email: m.frohme@dkfz-heidelberg.de.

FEATURES

location/Qualifiers

1..418

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="12"

/clone_lib="3.3 (upregulated in larynx carcinoma)"

BASE COUNT 71 a 149 c 120 g 78 t

ORIGIN

alignment_scores:

Quality: 767.00 Length: 139

Ratio: 5.518 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.281

alignment_block:

US-09-938-330-20 x AJ403134 ..

Align seg 1/1 to: AJ403134 from: 1 to: 418

2 ATCGAGCGGGGACCCCTCACGACGCTGAGGAGGCTGTCTGCTGGGC 51
 339 ahisSerGlnGlnArgGlnAspProSerHisAlaGlnHisHisAspHis 356
 52 ACATCCAGCAGCGCCAGGAGACCCAGCAGCGCTGAGACCATACACAG 101
 356 aValPheLeuThrArgGlnAspPheGlyProSerGlyTYrAlaProVal 372
 102 TTGTCTCTCTCACCCGCGGACCTTGGGCTTCAGGATGATGACCCGCTC 151
 373 ThrGlyMetCysHisProLeuArgSerCysAlaLeuasnHisGlnuAspG 389
 152 ACTGCAATGTCACACCCCTTACAGAGCTGTGCTCAACCATGAGATGG 201
 389 yPheSerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 406
 202 CTTCCTCTCACCTTCGATAGCTCATGAGACCGGCGGCTGCTGGCA 251
 406 etGlnHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuGly 422
 252 TGAGCATGATGATGATGAGGAGATGCTGTGATGATGATGATGATGATG 301
 423 SerValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpSe 439
 302 AGGCTATGAGCGGCGCCCTGCTGTCAGGCTGCTTCCACCGCTTCATTGGTC 351
 439 TArgCysSerLeuGlnGlnuSerArgTYrLeuProSerTYrAspCysAl 456
 352 CCGCTGAGCAGCAGCTGAGGAGCTCAGCGCTACCTCCCTCTACGATGGC 401
 456 leuAspAspProPhe 461
 402 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 418

seq_name: gb_est2:BJ096904

seq_documentation_block:

LOCUS BJ096904 555 bp mRNA linear (EST 12-DEC-2001)

DEFINITION BJ096904 NIBB Mochi normalized Xenopus early gastrula library

ACCESSION BJ096904

VERSION BJ096904.1 GI:17598016

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 555) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Kitayama, A., Terasaka, C., Mochi, M., Ueno, N., Shin-i, T. and Kohara

Y.

Expressed genes in X. laevis embryo

Unpublished (2001)

Contact: Tadasi Shin-i

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tsuhin@genes.nig.ac.jp.

FEATURES

location/Qualifiers

1..555

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="X15118"

/clone_lib="NIBB Mochi normalized Xenopus early gastrula

library"

/tissue_type="whole embryo"

/dev_stage="stage 10.5"

BASE COUNT 159 a 120 c 131 g 145 t

ORIGIN

alignment_scores:

Quality:	748.00	Length:	142
Ratio:	5.582	Gaps:	0
Percent Similarity:	94.366	Percent Identity:	84.507

alignment_block:

US-09-938-330-20 x BJ096904 ..

Align seg 1/1 to: BJ096904 from: 1 to: 555

```
452 SerTyrAspCysLeuLeuAspProPheAspProAlaTrpProGlnPr 468
|||||
129 TCATATGACTGCCTTCTGTGATGACCCCTTCGATCACAAATGGCCCAATT 178
|||||
468 oProGluLeuProGlyIleAsnTyrSerMetAspGluGlnCysArgPhea 485
|||||
179 ACCAGAACTCCAGGAATCATATTATCATGGATGACAAATGAGATTGCG 228
|||||
485 sPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheGluPro 501
|||||
229 ATTTCGTTGGGGCTACAGATGTGCACCTCTTCAGACTTTTGACCCCT 278
|||||
502 CysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCysLysTh 518
|||||
279 TGCAGAGCAGTTGTGTGACAGCCAGATATATCTTATTTCGCAAAAC 328
|||||
518 rLysLysGlyProProLeuAspGlyThrGluCysAlaProGlyLysTrpC 535
|||||
329 GAAGAAAGGGCCCTCCACTGTATGAGACAGAAATGCTCCCGAAGAGTGGT 378
|||||
535 yspheLysGlyHisCysIleTrpLysSerProGluGlnThrTyrGlyGln 551
|||||
379 GCTTTAAAGGATATTGCATTGTGAAGACACCAAGAGAGTCTTACAGTCAG 428
|||||
552 AspGlyGlyTrpSerSerTrpThrLysPheGlySerCysSerArgSerCy 568
|||||
429 GATGAGAGCTGAGACTCATGTGTCAAGTTTGATCTTGCTCGAACCTG 478
|||||
568 sGlyGlyGlyValArgSerArgSerArgSerCysAsnAsnProSerProA 585
|||||
479 TGGAGGTGAGAGTGCCTCCCGTAGACAGACAATGCAATTAACCCAGAGCAT 528
|||||
585 laTyrGlyGlyArgProCysLeuGly 593
:::|||||
529 CATATGGAGAGCGGATTGCCCTGCT 554
```


137 ...GlnGluaspPheargGluPheargGlnProLeuargGlnLucy 152
 427 CAGCCCGATGCACTGGGTGGTACGTAGGAGGACACACTTGGAGAACTG 476
 152 sValTyrThrGlyGlyValThrGlyMetProGlyAlaAlaAlaAlaIles 169
 477 CTGCTACAGGAGAGAGAGTGGGGATATGCAGAGCTCCTGGGTGCATCT 526
 169 eraPncysaspGlyLeuAlaGlyLeu...IleArgThrAspSerThrAsp 184
 527 GCACCTCTCTGGGCTCAGAGGCTTGGTGTCTGCACCCAGAGAGAGAGC 576
 185 PhePheIleGlu.....ProLeuGluArg 192
 577 TATACCTCGAGAGGAGGCGCTGGGACCTTCAGGCTCCTCCATTAATTC 626
 192 gGlyGlnGlnGlnGlyLeuAlaSerGlyArgThrHisValValTyrArg 209
 627 GCGATCCAGATCTCCACCTGCAGGCGCACACCTGTGCCCTGAGCTGCG 676
 209 rGlnGluAlaVal.....GlnGlnGlnTyrPalaGluProAspGlyAspLeu 223
 677 GGGATCTGTACACACTCAGACGCCACAGACGCCCTGGGACAGCGC 726
 224 HisaspGlnAlaPheGlyLeuGlyAspLeuProAsnLeuGlyLeuVa 240
 727 CAC..... 729
 240 lGlyAspGlnLeuGlyAspThrGluArgLysArgHisAlaLysProG 257
 730ATTCCGCGAGGCGGGATGTGTAACAG 757
 257 LysSerTyrSerIleGluValLeuLeuValValAspAspSer...ValVal 272
 758 AGACCAACACTGTGAGATTGTGTGCTGATCTCGAGAGGCCAG 807
 273 ArgPheHisGlyLysGlnHisValGlnAsnTyrValLeuThrLeuMetAs 289
 808 AAATACCGGAGCTTCCAGCAGCTGTAAACCGCACACTGGAAGTGGCCCT 857
 289 nIleValAspGluIleTyrHisAspGluSerLeuGlyValHisIleAsnI 306
 858 CTTCGCTGACATCTTCCGCGCC.....CTGATGTACAG 895
 306 lAlaLeuValArgLeuIleMetValGlyTyrArgGlnSerLeuSerLeu 322
 896 TGGCACTAGTGGCTGGAGGCTGACCCAGCCAGCTGACCTGGTGAATC 945
 323 lIleGluArgGlyAsnProSerArgSerLeuGlnGlnValCysArgTyrP 338
 946AGCCCAACCCAGCTGTACCTCGAATACTCTCCACTGGCG 989
 339 ...AlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHisLysA 354
 990 CAGGGCACTTGCCTGCTGCTGATGCC.....CATG 1021
 354 sPheIleValAlaPheLeuThrArgGlnAspPhe...GlyProSer...Gly 368
 1022 ACAGTCCCGACGTGGTACTGTACTTCTCTGGGCTCAGGCTGAGGC 1071
 369 TyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaLeuAs 385
 1072 ATGGCATTTCAGAACTCATCTGTCTCTGACTTCTCAGAGAGTGTGAA 1121
 385 nHisGluAspGlyPheSer.....SerAlaPheValIleAlaHisG 399
 1122 CATGACCACTCCACCGCATCTCGGAGTCCCTCTCCATATGCCCATG 1171
 399 luthrGlyHisValLeuGlyMetGlnHisAspGlyGlnGlyAsnGlyCys 415
 1172 AGTTGGGCGCACGCTGGGCTGTGACCATGATTTCCTGGGAATAGCTGC 1221

416 AlaAspGlnThrSerLeuGlySerValMetAlaProLeuValGlnAlaI 432
 1222CCCTGTCCAGGCTCCAGC 1238
 432 aPheHisArgPheHisIleTrpSerArgCysSerLysLeuGluLeuSerArgT 449
 1239 C..... 1239
 449 TyrLeuProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTyr 465
 1240CCAGCCAAAGACTGCATCATCATGAGAGCCCTCCACAGAC..... 1275
 466 ProGlnProProGluLeuProGlyIleAsnTyrSer..... 477
 1276TTCCTACAGGCTCGAATCTCACCACTGACGAGCCGACG 1313
 478 .MetAspGlnGlnCysArgPheAspPheGlySerGlyTyrGlnThrCysL 494
 1314 GGCCCTGGAGAAAGCCCTCTGTGAT.....GGAATGGGCGACCTGCC 1354
 494 euAlaPheArgThr..... 498
 1355 TCTTCGACAGGCTGCTACCTACCCCTATAGCTGCTTTCGCGGAAT 1404
 499PheGluProCysLysGlnLeuTyrCysSerHisProAspAsnPr 513
 1405 ATGTTGTGAGAGCGGGCGAGCATGTGACTGTGCTTCTTCCTGATGAC.. 1452
 513 cTyrPheCysLysThrLysGlyGlyProProLeuAspGlyThrGlyCys. 529
 1453TGC.....GTGATCTCTGCTGTGATTTCTTGACCTGCC 1486
 530AlaProGlyLysTyrCysPheLys...GlyHisCysIleTyrLys 543
 1487 AGCTGAGGCCAGGTGCACATGTGCACTGACAGGACCTGTGT..... 1530
 544 SerProGlnGlnThrTyrGlyGlnAspGlyGlyTyrTrpSerSerThrLys 560
 1531CAAAATTCAGACGTCGCCCGCTGCTGAGTGCAGATGCTGCTAC 1574
 560 sPheGlySerCysSer.....ArgSerCysGlyGlyAlaValArgSera 575
 1575 CAGAGGGAGATGTGACTTGCCTGAATTCGCCACAGAGAC..... 1614
 575 rGSerArgSerCysAsnAsnProSerProAlaTyrGly...GlyArgPro 590
 1615 ...AGCTCCAGTGT.....CCCCGTGATGTCAAGCTTACGGGATGCGAGGCC 1659
 591 CysLeuGlyProMetPheGluTyrGlnValCysAsnSerGlnGluCysPr 607
 1660 TGGCCTGGCGGCA.....GCTGTGTGATGCAGCGGCGTGT... 1698
 607 oGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgAsnSerT 624
 1699 .GCCTCTAT.....GCCAGCAGTGC..... 1719
 624 yTyrValHisGlnAsnAlaLysHisSerTyrValProTyrGluProAsp 640
 1720CAGTCACTTGGGAGCT..... 1737
 641 AspAspAlaGlnCysGlnLeuIleCys...GlnSerAlaAspThrGln 656
 1738 ...GGAGCCAGCGCGCTGGCGCACTTGGCTCCAGACACTAATACTCG 1784
 656 yAspValValPheMetAsnGlnValValHisAspGlyThrArgCysSerT 673
 1785 GGAAATGCTTT..... 1797
 673 yTyrArgAspProTyrSerValCysAlaArgGlyGlyCysValProValGly 689
 1798GGGACTGT.....GGG 1809
 690 CysAspLysGlnValGlySerMetLysAlaAspAspLysCysGlyAlaCy 706

```

1810 CGCAACCCAGTGGAGTTATGTCTCTGACCCCTAGAGATGCAATTTG 1859
706 sgllylaspasnserhiscysargthrvalylsglythrleuglylva 723
1860 TGGG.....CAGCTCCAGTGCCAGACA.....GGTAGGA 1888
723 laserlysglnalaglalaLeuValGlnleProAlaGlyAla 739
1889 CCCAGCTCTGCTGGGCTCCATC..... 1911
740 ArgHislleGlnleGlnAlaLeuGlyLysSerProHlsArgSerValva 756
1912 CGGAGATCTACTCTGGAGACAATAGAT..... 1938
756 LysAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGlnua 773
1939 .....GTGAATGG..... 1947
773 laThrSerArgThrPheThrAlaMetGlyLeuGluTrpGlu.....Asp 787
1948 .....ACTGAGCTGAACTGCAGCTGGGTGCACCTGAC 1980
788 AlaValGluAspAlaLysGlnSerLeuLysThrSerGlyProLeuProG1 804
1981 CTGGCGAGTGTGTGGCCCGCCCTC..... 2007
804 uAlaIleAlaIleLeuAlaLeuProProThrcGlnGlyGlyProArgSers 821
2008 .....CTGACTCTCTCCGCGACAGACCTGTGGCCCT..... 2037
821 eLeuAlaIlyrLysTyValIleHlsGlu.....AspLeuLeuProLeuIle 836
2038 .....GGCCTGTGTGTATAGACATCATGATGACCGCTGCATCTCCTG 2082
837 GlySerAsnAsnValLeuLeuGlnLumetAspThrTyrgluTrpAlaLe 853
2083 GGGCGACAGAA..... 2094
853 uLysSerTrpAlaProCysSerLysAlaCys...GlyGlyGlyIleGlnP 869
2095 .....TGTGAGAGCAATGCCATGAGCATGGGCTCTGCTG 2128
869 heThrLysTyrglyCys ArgArgArgArgAspHlsMetValGlnar 885
2129 ACAGAAAGAGGACGCTGACTGTAGGAGGCTGGCAGACCCCTGACTGC 2178
885 gHlsLeuCysAspHlsLysLysArgProLysProIleArgArgArgCysa 902
2179 ACCACTCAGCTCAAGACACGCTCCCTGACCA..... 2212
902 snGlnHlsProCysSerGlnProValTrpValThrGlnGluTrpGlyAla 918
2213 .....CAGGGCTGCTCCACGCTCTGCTGTATGTGCTCTG..... 2251
919 CysSerArgSerCysGlyLysLeuGlyValGlnThrArgGlyIleGlnC 935
2252 .....TGATG 2256
935 sLeuLeuProLeuSerAsnGlyThrHlsLysValMetProAlaLysAlaC 952
2257 CTGTGTGCA...CTACTGCTGCTAC.....GTCCCGGCT 2288
952 ysAlaGlyAspArgProGlnAlaArgArgProCysLeuArgValProCys 968
2289 GCACGAGGACTCTGCCAGCTCAAGGACCA..... 2320
969 ProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysgl 985
2321 CCTGCC.....AGTACAGGCGACACCAATCTGCTCCTCT 2355
985 yGluGlyTlIeGlnGlnArgGlnValValCysArgThrAsnAlaAsnSerL 1002

```

```

2356 GAACGGCAGGACCTCCGAGAGGGCTCTGTCGACAGGACGACTAAAGC 2405
1002 euGlnHiscysGlnGlyAspArgProAspThrValGlnValCysSerLeu 1018
2406 TA...GTGCTCTCAGCTCCCGGCCCCCTTCAGAGCGCGCTGCCCTG 2452
1019 ProAlaCysGlyGlyAsnHlsGlnAsnSerThrValArgAlaAspValTr 1035
2453 ACCCTGTGT.....CCAGAGACTCCAGTCTCAG..... 2482
1035 pGluLeuGlyThrProGlnGlyGlnTrpValProGlnSerGlyProLeuH 1052
2483 .....GCCCAAGCAAGCC.....C 2498
1052 lsProIleAsnLysIleSerSerMetCysAlaAlaGluProCysThrGly 1068
2499 ACCCC.....CAAGGAGCACTGCTGCTGCCAGCCCAAGGCGCGGT 2539
1069 AspArgSerValPheCysGlnMetGluValLeuAspArgTyrcysSeril 1085
2540 GCCCATCGGTGACCTGCCCGGCGCAGGGCGTGAA.....TCCCG 2580
1085 eProGlyTyrcys.....HisArgLeuCysCysValSerc 1096
2581 CCCCTAGTGTGACCTCCAGACAGCGCCACCGCTC.....CGACAGT 2624
1096 ysIleLysLysAlaSerGlyProAsnProGlyProAspProGlyProThr 1112
2625 GTCTCGCTCTACTCTGACCTCGGAGGT..... 2656
1113 SerLeuProProPheSerThrProGlySerProLeuProGlyProGlnAs 1129
2657 .....TCCGCTGCTCCAAAGCGGACTTAAAGGCTTAA.. 2689
1129 pProAlaAspAlaAlaGluProGlyLysProThrGlySerGlnAsp 1145
2690 .....GAGCGGCGTGGCCCTCGAGTGCCTTACATGACGAAAGC 2732
1146 HlsGlnHlsGlyArgAlaThrGlnLeuProGlyAlaLeuAspThrSerc 1162
2733 GCCAGAGCTGGCGGTGTAAAGATCCGGC.....ACCGCAC 2773
1162 rProGlyThrGlnHlsProPheAlaProGluThrProIleProGlyAlas 1179
2774 GCGCTGTAAAGCAACACTCTCGGACCTGCCCGGCTGTGCAAGCGGGG 2823
1179 eTrpSerIleSerProThrThrProGlyLysLeuProTrpGlyTrpThr 1195
2824 CTTCG.....GAGGGGCTGGGGGTGGAGC 2849

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMP.seq:US-09-920-048-1

```

seq_documentation_block:
? Sequence 1, Application US/09920048
? Patent No. 6344352
? GENERAL INFORMATION:
? APPLICANT: MERKULOV, Gennady et al
? TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
? TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
? FILE REFERENCE: CL001177DIY
? CURRENT APPLICATION NUMBER: US/09/920,048
? PRIOR APPLICATION NUMBER: 09/813,819
? PRIOR FILING DATE: 2001-03-22
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 1
? LENGTH: 2968
? TYPE: DNA
? ORGANISM: Human
US-09-920-048-1

```

alignment_scores: Length: 1268
 Quality: 310.00 Gaps: 73
 Ratio: 0.598
 Percent Similarity: 40.852 Percent Identity: 22.082

alignment_block:
 US-09-938-330-20 x US-09-920-048-1 ..

Align seg 1/1 to: US-09-920-048-1 from: 1 to: 2968

```

3  ProLeuArgAlaLeuSerTyrLeuLeuProLeuHisCysAlaLeu.. 18
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55  CCGGCCGGAGCGCGCGCTGCCATGCGGCTGCGCTCTGGGCGCTGGG 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19  CysThrAlaAlaGlySerArgThrProGluLeuHisLeuSerGlyLys 35
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 GCTCTGGGCGCGCGCGCGCTGCCATGCGGCTCTGGGCGCTGCCAATATAG 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35  euSerAspTyrGlyValThrValProCysSerThrAspPheArgGlyArg 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 CCGTG.....CTGCGATGCC.....TCA 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52  PheLeuSerHisValSerGlyProAlaAlaSerAlaGlySerMe 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 GTACTGCTTGGGGTGCTCTGGGACCTGCAGTGGC..... 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68  tValValAspThrProThrLeuProArgHisSerSerHisLeuArgV 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
211 .....ACTGAGAGCAGCAGGCGAGCAGTCCAGAGAGAG 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85  aAlaArgSerProLeuHisProGly..... 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 CCCGAGGAGCGCTTGGAGCCCGCAGTCTTCCAGAGCATCTCCCAT 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94  .....GlyThrLeuThrProGlyArgValGlyArgHis 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
292 ACCCTCAAAAAGCTGCTCAGACAGCTGCCT.....GA 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104  sSerLeuTyrPheAsnValThrValPheGlyLysGluLeuHisLeuArgL 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 GCCCTGAGAGTCAAGTGGAGCTGAGCGGTACAGTATCTCTGGAGC 376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121  euArgProAsnArgArgLeuValValProGlySerSerValGluTrp... 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377 TGTCTACAGATATAGGAGTGTGTCCAGCGCGCCAAACCTGTGTGGTAC 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137  ...GlnGluAspPheArgGlyLeuPheArgGlnProLeuArgGlnGly 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
427 CAGCCGATGGCAGCTCGGCTGCTCAGTGGAGGACACACTTTGGAGACTG 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152  sValTyrThrGlyGlyValThrGlyMetProGlyValAlaAlaAlaHis 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 CTGCTACACAGGAGAGTGGGAGATATCCAGGCTCTGGGTGTCTCAT 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169  eArgAsnCysAspGlyLeuAlaGlyLeu...IleArgThrAspSerThrAsp 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 GCACCTGCTGTGGCTCAGAGGCTTGTGTCTGTACGCCCAAGAGAAAC 576
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185  PhePheIleGlu.....ProLeuGlnArg 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
577 TATACCTGTGAGCAGCGGCTGGGACCTTCCAGGCTCTCCCATATATTC 626
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192  gGlyGlnGlnGlnLysGluAlaSerGlyArgThrHisValValTyrArg 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
627 GCGAATCCAGAGATCTCCACTGCCAGGCGCACACTGTGCTGAGCTGAG 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209  rGlnLysAlaVal.....GlnGlnGlnLysLysLysLysLysLysLys 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 GGAATCTGTACACTCAGAGCGCACAGAGACCCCTGGGAGACGCGC 726
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224  HisAsnGlnLysLysLysLysLysLysLysLysLysLysLysLys 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

727  CAC..... 729
240  tGlyAspGlnLeuGlyAspThrGluArgLysArgHisAlaLysProG 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
730  .....ATTGCGGAGCGGCGGATGGTACAG 757
257  LysTyrSerIleGlnValLeuValAlaAspSer...ValVal 272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
758  AGACCAAGACTGTGAGTGTGATGTGGCGATCTCGAGGCGCCAG 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
273  ArgPheHisGlyLysGlnHisValGlnAsnTyrValIleThrLeuMet 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
808  AAATACCGGAGCTTCCAGCAGCTGTAAACCGCACAGGAGTGGCCT 857
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289  nIleValAspGluIleTyrHisAspGluSerLeuGlyValHisIleAsn 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
858  CTGTGCGACACTTCTTCCGCGC.....CTGAATGTACAG 895
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306  leAlaLeuValArgLeuIleMetValGlyTyrArgLysSerLeuSer 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
896  TGGCACTAGTGGGCTGAGCGCTGAGCCAGCGAGCTGCTGGAGATC 945
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323  tIleGluArgGlyAsnProSerArgSerLeuGlnGlnValCysArgTrp.. 338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
946  ....AGCCCAAAACCGAGCTGTACCTCGAATACTTCTTCACATGGCG 989
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339  ....AlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHisHis 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
990  CAGGCGACATTTGCTGCTCGATTGCC.....CARG 1021
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354  sPheValValPheLeuThrArgGlnAspPhe...GlyProSer...Gly 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022  ACAGTGCCGAGCTGTGAGTGTACTTCACTTCTCTGCGCTGAGTGGCG 1071
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369  TyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaLeuAs 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1072  ATGGCCATTAGAACATCCATCTTCTCCTGCACTTCTCAGAGGTGTGA 1121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385  nHisGluAspGlyPheSer.....SerAlaPheValIleAlaHisG 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122  CATGAGCAGCTCCACAGCATCTCTGGAGTGCCTCTCCATAGCCCATG 1171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399  tLutThrGlnHisValLeuGlyMetGlnHisAspGlyGlnGlnGlnGly 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1172  AGTTGGCCACAGCTGTGGCTGAGCATATTTGCTGGGAATTAATCTG 1221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416  AlaAspGluThrSerLeuGlySerValMetAlaProLeuValGlnAla 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1222  .....CCGTGTCAGGTCACG 1238
432  aPheHisArgPheHisTrpSerArgCysSerLysLeuGluLeuSerArgT 449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1239  C..... 1239
449  yIleuProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTrp 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1240  ....CCAGCCAAAGACCTGCATCTGAGGCTCCCAAGAC..... 1275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
466  ProGlnProProGluLeuProGlyIleAsnTyrSer..... 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1276  .....TTCCTACAGCGGCTGAGCTTCACTGCAACTGCAGCGCAGC 1313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478  MetAspGlnGlnCysArgPheAspPheGlySerGlyTyrGlnThrCysL 494
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1314  GGCCTGTGAGAAAGCCTCTGTGAT.....GGAATGGCAGCTGCC 1354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
494  euAlaPheArgThr..... 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1335  TCTTTCGAACGAGCTGCTAGCTACCCCTAGTGTCTTCTGCGGAAT 1404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499  ....PheGluProCysLysGlnLeuTrpCysSerHisProAspAsnPr 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405  ATGTTTGTGAGCGGCGGAGCAGTGTGAGTGTGCTTCTCTGATGAC.. 1452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

513 OTYrPheGysLysThrLysLysGlyProPLeuAspGlyIthrGlyCys. 529
1453TGC.....GTGCATCCCTGCTGTGATCTTTTGACCTGCC 1468
530AlaProGlyLysTrpCysPheLys..GlyIstysLleTrpLys 543
1467 AGCTAGAGCCACAGGTGCACACTGTGCATCTGCAGGACCCGTGTT. 1530
544 SerProGluGlnIthrTrgIrgLysAspGlyLeuIrgPheSerSerTrpThrLy 560
1531CAAAATTGCCAGTGCGCCCGCTGTGGGTGCAGTGTGCCTGCAC 1574
560 sPheGlySerCysSer.....ArgSerGysLysGlyValArgSerA 575
575 CAGAGGGGATGTACTTGCTTCAATTGCTGCCAGCAGAC..... 1614
575 rGSerArgSerCysAsnAsnProSerProAlaTrgLy..GlyArgPro 590
1615 ..ACCTCCCACTGT..CCCCGTGATGTACGCTAGGGATGGCAGGCC 1659
591 CysLeuGlyProMetPheGlyIthrGlnValCysAsnSerGluGlyCysPr 607
1660 TGCCCTGGCGGCA.....GCTGTGTCAAGCAGCGGCGGTGT.. 1698
607 oGlyIthrTrgGluAspPheArgAlaGlyAlaLysArgAsnSer 624
1699 ..GCTCTAT.....GCCAGCAGTGC..... 1719
624 yTrgValIhrGlnAsnAlaLysHisSerTrpValProTrgIhrProAsp 640
1720CAGTCACTTTGGGGACT..... 1737
641 AspAspAlaGlnCysCysGluLeuIecys..GlnSerAlaAspTrnG 656
1738 ...GAGCCACGCCCTGCGCCACTTTGCTCCAGACGTAATACTCG 1784
656 yAspValValPheMetAsnGlnValValHisAspLyThrArgCysSer 673
1785 GGGAAATGCTTT..... 1797
673 yTrArgAspProTrgSerValCysAlaArgGlyGlyCysValProValGly 689
1798GGACCTGT.....GGC 1809
690 CysAspLysGluValGlySerMetLysAlaAspAspLysCysGlyValCy 706
810 CGCAACCCAGTGGCAGTTATGTCTCGACCCCTAGAGTGCATTGG 1859
706 sGlyGlyAspAsnSerHisCysArgTrhValLysGlyThrLeuGlyLysA 723
1860 TGGG.....CAGCTCCAGTGGCACA.....GTAAGA 1888
723 lAsrLysGlnAlaGlyAlaLeuLysLeuValGlnIleProAlaGlyAla 739
1889 CCCAGCCTGTCTGGCTGCATC..... 1911
740 ArgHisIleGlnIleGlnAlaLeuGlyLysSerProHisArgSerValVa 756
1912 CGGATCTACTCTGGGAACAATAGT..... 1938
756 lLysAsnGlnValTrgIrgLysSerPheIleLeuAsnProLysGlyLysGln 773
1939GTGATGG..... 1947
773 lathSerArgTrhPheThrAlaMetGlyLeuGlyTrpGlu.....Asp 787
1948ACTAGCTGAACGTGACGTGTGGTCACCTGGAC 1980
788 AlaValGluAspAlaLysGlySerLeuLysThrSerGlyProLeuProL 804
1981 CTTGGCAGTGTGTGCCACCCCTC..... 2007

[illegible]

```

1113 SerLeuProProhSerThrProGlySerProLeuProGlyProGlnAs 1129
2657 .....TCCGCTGCCTCCAGCCGCGACTTAAAGGCTTCA.. 2689
1129 proAlaAspAlaAlaGluProProGly LysProThrGlySerGluAsp 1145
2690 .....GAGCGGGGGCTGCTGAGAGTCCCTTACATGACTGAAAGCC 2732
1146 HisGlnHisGlyAlaGluAlaThrGlnLeuProGlyAlaLeuAspThrSer 1162
2733 GCCAGAGACTGGGGGTCTTAAAGACTCCGGG.....ACGGCCAC 2773
1162 rProGlyThrGlnHisProhLeuProGlyThrProLeuProGlyAlaAs 1179
2774 GCCCTGTCAAGCAACACTGCGGAGCTGCGGCGTGGTGCAGCGGGG 2823
1179 eTrpSerIleSerProThrThrProGlyGlyLeuProThrGlyThr 1195
824 CTGG.....GAGGGGCTGGGGGTGGAGC 2849
seq_name: /cgn2.6/ptodata/2/ina/6A.COMB.seq: US-08-985-526-4

```

```

seq_documentation_block:
; Sequence 4, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; NUMBER OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Boyle, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1326 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-985-526-4

```

```

alignment_scores:
  Quality: 297.00      Length: 587
  Ratio: 1.356        Gaps: 24
  Percent Similarity: 37.308  Percent Identity: 20.443
alignment_block:

```

```

US-09-938-330-20 x US-08-985-526-4
Align seg 1/1 to: US-08-985-526-4 from: 1 to: 1326

```

```

473 GlyIleAsnTyrSerMetAspGluGlnCysArgPheAspPheIleArg 489
64 GAGTGTACATACAGAAATTAACAGAGAAATGAGATGTTCAT.....AG 104
489 yTyrGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeu 506
105 CTGCACAGAGTGTACTGTCTCAGAACTAGTACCATGTGCAAAAGGTGT 154
506 rPcysSerHisProAspAsnProTyrPheCysLysThrLysGlyPro 522
155 CCTGC.....CCCATCATGCC.....TGCATCAATGCGACAGTTCT 192
523 ProLeuAspGlyThrGluCysAlaProGlyLysThr..... 534
193 .....GATGGA...GAATGCTGTCTGCTGTGTGGCCACAGCACTTGC 233
535 ..... 535
234 GAGCAGATGGCTGCTGCTCATATGTCGAGTGCAGCTCTCTGTTCAGAGCT 283
535 yspPheLysGly..... 538
284 GTGGCAATGATTAATTCAGACGCGCGCGCTCTGATAGCTCAACAC 333
539 .....HisCysIleTrpLysSerPr 545
334 CGATGTGAGGGCTCTCGTCCAGACAGCAGCTGACCATTCAGAGAGTG 383
545 oGluInThrTyrGlnAspGlyLysPheSerThrLysPheG 562
384 TGACAAAGATTTAAACAGAGATGCTGAGCGACAGTCCCGCGTGT 433
562 ySerCysSerArgSerCysGlyGlyValArgSerArgSerArgSer 578
434 CATCTTGTCTGTGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 483
579 CysAsnAsnProSerProAlaTyrGlyAlaArgProCysLeuGlyPro 595
484 TGCACCTCTCCAGCCGCCAGATGATGGAACCTTGAAGGAGAGC 533
595 tPheGlyTyrGlnValCysAsnSerGluGluCysProGlyThrTyrGlu 612
534 GCGGGAGACCAAGCCTGCAAGAAAGACGCTGCC..... 570
612 spPheArgAlaGlnGlnCysAlaLysArgAsnSerTyrTyrValHisGln 628
570 ..... 570
629 AsnAlaLysHisSerTrpAlaProTyrGluProAspAsp..... 641
571 ...ATCAATGAGAGGCTGGGGCTCTGTGATCAGATGTGTGTGT 617
642 .....AspAlaGlnLysCysGluLeuIleCysGlnSerAla. 653
618 CACCTGTGAGAGAGGGGTACAGAAAGTGTGTGTGTGTGTGTGTGT 667
654 ....AspThrGlyAspValAlaPheMetAsnGlnVal..... 664
668 GAATGACTAGAGAGCAAGAGTTGGCCATATGAGCTGAGGGGCTGCC 717
665 .....ValHisAspGlyThrArgCysSerTyrArgAspProTyrSer 679
718 CTATGCTATCAGACAGAGAGTTCAGTACAGAAATTAACAGAGATGAGCT 767
679 lCysAlaArgGlyGluCysValProValAlaLysCysAspLysGluValGlys 696
768 T..... 768
696 erMetLysAlaAspAspLysCysGlyValCysGlyLysAspAsnSerHis 712

```

```
seq_name:cgn2_6/plodata/2/ina/6A_COMB.seq;US-08-706-216-5
seq_documentation_block:
? Sequence 5, Application US/08706216
? Patent No. 6140098
? GENERAL INFORMATION:
? APPLICANT: Balasubramanian, Sriram
? APPLICANT: Ford, John
? APPLICANT: Gorman, Daniel M.
? APPLICANT: Zurawski, Gerard
? TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: DNAX Research Institute
? STREET: 901 California Avenue
? CITY: Palo Alto
? STATE: California
? COUNTRY: USA
? ZIP: 94304-1104
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/706,216
? FILING DATE: 30-AUG-1996
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Ching, Edwin P.
? REGISTRATION NUMBER: 34,090
? REFERENCE/DOCKET NUMBER: DX0613
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-852-9196
? TELEFAX: 415-496-1200
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1173 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1173
? US-08-706-216-5
alignment_scores:
? Quality: 293.00 Length: 407
? Ratio: 1.363 Gaps: 21
? Percent Similarity: 52.826 Percent Identity: 29.238
alignment_block:
US-09-938-330-20 x US-08-706-216-5 ..
Align seg 1/1 to: US-08-706-216-5 from: 1 to: 1173
76 LeuPRArghISerSerhNISleAuGvAlAlArGSe...ProLeuNI 91
||||| ||||| ||||| |||||
97 TTGGCGGCTGTCTGCACGAGCCGCCAGTCCGCCGAGCTCCTCCCA 146
||||| ||||| ||||| |||||
91 sPrGIGLyThLeuTrPrGvIyArGvAlGvIArghISerLeuTyR 108
||| ||| ||||| |||||
147 CTTCGGC.....CTGCACCACAGAGGGGTGAC.....T 175
||||| ||||| ||||| |||||
108 heAnVAIrPhaIphagILySGvIeUnISleAUgLeuArGrPrASn 124
::: :::: ||| :::: |||:::||||| |||
176 AGNCSTTGAGGCACAGGGAACAATTCAACCTCACCTCGGAGAAGAC 225
||||| ||||| ||||| |||||
125 ArGArGLeuVAlValPrGvIySerSerValGIuTyR...GInGIAsPrh 140
```


9201 TCACCCAGCGCGTCCCGTACCGTGGCGCTCCACACATCC..... 9160
 52 leuSerHisValValSerGlyProAlaIleSerAlaGly_SerMet 68
 9159 ...ACGTCGGGGGACACCGCCGATCGACGAGCGCGCTCACGGA... 9117
 69 ValValAspThrProThrLeuProAlaHisSerSer_HisLeuArg 85
 9116TCACCGCGTGGCGCGGACCACTCGCGCGCTGACGCCA 9076
 85 AlaIleArgSerProLeuHisProGlyGlyThrLeuTrpProGlyArgVal 101
 9075 TTGCTGCG.....ACCGTCTGATTGACCGGCACT 9047
 102 GlyArgHisSerLeuTyrPheAsnValThrVal.....PheGly 115
 9046 GCCCGCGCACCGCCGACGACCGCATACCGTTGCGCGCGCGCTCGGAAA 8997
 115 sGluLeuHisLeuArg.....LeuArgProAsnArgArgLeu 128
 8996 GCCGCTCCACCGACGACACCCACACCTCCGCCGACGCGCTCCGCTC. 8948
 128 AlaValProGlySerSerValGluTrpIleGlnAspPheArgGluLeuPhe 144
 8947CGACCCCTCCGCGAAGCGCTT 8927
 145 ArgGlnProLeuArgGlnGlyCysValTyrThrGlyVal_ThrGlyW 161
 8926 GCACCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8877
 161 eTrpGlyAlaIleValAlaIleSerAsnGlyAspGlyLeuAlaGlyLeu 177
 8876 CCGTGGCG 8849
 178 IleArgThrAspSerThrAspPheIleGluProLeuGluArg..... 192
 8848ACCGCCACAGCCGATGACATCACCCCGCGCGCGCGCGCGCGCG 8805
 193 GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 207
 8804 CCGCGCATGTCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8755
 208 ArgArgGlnAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 222
 8754 AGCG 8705
 223LeuHisAsnGlnAlaPheGlyLeuGlyAspLe 233
 8704 ACTGCCGAGGCGTCCCGTGAGCACATGCCCGCGTGTGTGCGCGCGCT. 8655
 233 uProAsnLeu.....LeuGlyLeuValGlyAspGlnLeuGlyAspThr. 247
 8654 CCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 8605
 248GluArgGlyAspArgHisAlaGlyAspGlySerTyr 259
 8604 GTCCGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8564
 260 SerIleGluValLeu.....LeuValValAspAspSerValValArgPh 274
 8563 ...CTCCAGCGGCTCCGACGAGGTCTCCGACGACGACGAGCGGTGGTGGTGG 8517
 274 eHisGlyGlyGlnHisValGlnAsnTyrValLeuThrLeuMetAsnIleV 291
 8516 CCATGGCGCGTGGCTCGCGCGCGCATAT..... 8499
 291 AlaAspIleTyrHisAspGlnSerLeuGlyValHisIleAsn..... 305
 8488CCGGAAGACTCGCGGTGAACCTCGCGCGCGCG 8456
 306IleAlaLeuValArgLeuIleMetVal 314
 8455 ACCGAGGACCGCGCTCCCGTACATACCTCTTCCCGCGCGCGCGCGCT 8406

314 IGlyTyrArgIleSerLeuSerLeu.....IleGluArg 326
 8405 CCGGCTCTACAGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8356
 326 LysAsn.....ProSerArgSerLeuGlnGlnValCysArgTyr 338
 8355 GAGACCGTGTCCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8306
 339 AlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHisAspH 355
 8305 CGCCACAC...GCCCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8259
 355 sValValPheLeuThrArgGlnAspPheGlyProSerGlyTyrAlaProV 372
 8258 GTCACGCTGTCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8222
 372 aThrGlyMetCysHisProLeuArgSerGlyAlaLeuAsnHisGluAsp 388
 8221CGTGGCGCGCTCCGCGCGCGCTGTC 8198
 389 GlyPheSerSerAlaPheValIle.....AlaHisGluThrG 396
 8197 GGG...ACCGACAGCTTCTGCTGTCAGGTGGAGCGGACGAGGCGG 8151
 397AlaHisGluThrG 401
 8150 TCGGCGTGTGAGAGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 8101
 401 LysHisValLeuGlyMetGlnHisAspGlyGlnGly_AsnGlyCysAlaAs 417
 8100 GTCAGCGCGGTACCGAGTTCGAGACGCGTGAAGGAGTGAACCGCGCGGA 8051
 417 P...GluThrSerLeuGlySerValMetAlaPro.....LeuV 429
 8050 CCGGAAGCGCGAGCTCCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCG 8001
 429 aGlnAlaAlaPheHisArgPheHisTrpSerArgCysSerLysLeuGlu 445
 8000 TCCGCGCGCGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7951
 446 Leu.....SerArgTyrLeuProSer...TyrAspC 455
 7950 GCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7901
 455 sLeuLeuAspAspProPheAspProAlaTrpProGlnProProGluLeu 472
 7900 T.....CCGCTCCCGCGCGCGTGCCTGCGCGCGCGCGCGCGCGCG 7863
 472 roGlyIleAsnTyrSerMetAspGlnGlnCysArgPheAspPheGlySer 488
 7862 CTGGGCGT.....AGAGCGCGACGCGTGC 7837
 489 GlyTyrGlnThrCysLeuAlaPheArgThrPheGlu...ProGlyS 503
 7836 GCGCGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7787
 504GlnLeuTrpCysSerHisProAspAsnProTyrPheCysL 517
 7786 CGACGAGTACGCGCGGTGACCGACCGCGCGCGCGCGCGCGCGCGCGCG 7737
 517 yThrLysLysGlyProProLeuAspGlyThrGlnCysAlaPro...GlyL 533
 7736 CTTGGCGCGAGCGAGGAGCTCTGCGCGCGCGCGCGCGCGCGCGCGCG 7687
 533 yTrpCysPheLysGlyHisCysIleTrpLysSerProGlnGlnThrTyr 549
 7686 GAGGTCCCG 7637
 550 GlyGlnAspGlyGlyTrpSerSerTrpThrLysPheGlySerSerArg 566
 7636 GGGGTCCAGGGC.....CGTACGCGCGTCCCGTTCAGAGGACTTCCGCG 7593

```

566 gsercysgllygllyvalargserarg..... 577
5792 ccgttccgagagccattccggacacggcccccagggcattcgag 7543
578 ..sercysasnproserprola..... 585
7542 cgggcacggcgtccctggccggccgctccgacggcgctcgagcgc 7493
586 ..tyrsglyarg..... 597
7492 ggcgttgccggccgctcgacggcctccgtccgttgcacca..... 7451
597 utyrghvalcysasnserglucysprogltyrtyrghuaspphea 614
7450 .cagccgcagatgagagaaagacagagcgctcgagcgaagctcc 7402
614 rgalaglnincysalalysarganserlyrtyrvalhisglnasna 630
7401 gcggtag..... 7394
631 lyhissertrypalprotyrghuaspaspapalagllyscysgl 647
7394 ..... 7394
647 uleuilecysglnseralaspthrlyspvalval..... 659
7393 ....cttcgtccagtgacggcgccgctgacgttgcgcgacgtacggcg 7348
660 .....phenetanglnalvalhisasrgly 668
7347 tccaccccgctcgagctgacggacgtcgacacacgcgtcgacacac 7298
669 thargcysertyrargasprotyrservalcysala..... 661
7297 tgcgcctgcagatgacggcgccgctcctcgcgagggcccgca 7248
682 ....argglglucysvalproval.....glycysasplysg 693
7447 ggaagcgccagtgctcccgctcgacgtcgacggcctcgacggcc 7198
693 luvalglysermetlysalasp.....asplyscysgllyvalcysgly 707
7197 acttcggacacgacggcgtcagctcctcgacgagttcgccactcgcg 7148
708 glyspasn.....serhisctysargthrvalysgllythrleu 721
7147 gccggccggccggccggccggccagagagagagcggcgcc.....gc 7104
721 ylusalaserlysglnalaglalauleuylsleuval..... 733
7103 ccgacccggcgagcctcgcgccgacgtcgacacccagggcaccggtgcc 7054
734 .....glnileuproalagly.....alaarghisilegln 743
7053 ccggtagatcagtagcgtgcccggagaccacggcgtcgccgacgacac 7004
744 ile.glnalaleuengllyserprohisargservalvallysasnly 760
7003 gtccagagtgctgcttcgctcctcgccacggcgctggcgggcgcg 6954
760 a.....thrghyserrheilleuasnpro 768
6953 tggtagcgtcgacggcgagcgccggagcgggagc..... 6919
769 lysglylysghualathrserargthrphenrhametgly.....le 783
6918 ..ggcggggagacacacggcgacatgacggcgagggcgctaccgc 6872
783 uglutrp..... 785
6871 gaactggttcttccccggcgagacggcccccagacgtacgagaca 6822
786 .....gluaspalaval.....gluaspalalys..... 793

```

```

6821 cccggcgccgacgactccggcgagctccgacagggcccccagggccgc 6772
794 ..gluserleuylthrserglyproleuproglnalaleallele 809
6771 ccaggttcagacggcgacgtcgatgacgtccacagctgacg..... 6727
809 ualaleupprothrleu.....glyproargserleualat 824
6726 .....ccccgaggaagggtctgctcgccggcgacggccgctc 6684
824 yrltyrvalilleisgluaspleuauproleuileglyserasn 840
6683 caccggtggtatccagac.....ggggttcgggt 6652
841 valleuenglumetaspthrtyrghlupralaleuylsertrp.. 856
6651 accgc.....tcgccagcgcctggac 6629
857 ....alaprocyserysalacysgllygly.ileglnpherhlys 871
6628 gaggcgacggcgtcgccgctcgccgctcgccggcgacgacagaa 6579
872 tyrglycysargargargargasphehismetvalglnarghisleucy 888
6578 ctgggtcctcgggcgccgacgctcac..... 6552
888 sasphistylsargpro..... 894
6551 .....ggcgcccccagacaggtcccgctcgtccggcgt 6512
895 ..lysprolleargargargcysanglnhisprocysserghproval 910
6511 accgacccgacggcgacgctgagctcgcccccggcgacggcgagt 6462
911 trpvalthrngluutrpelyalacyserysercysgllylsleu 927
6461 cactgcatcgggggcgagctcgagctcgccgctcctcgtccgg 6412
927 yval.....glnthargglylleghincysleu 938
6411 gccagagacacatgacggcgacggcggtggcg..... 6372
938 roleserasnlythrhislysalmetproalaleuylalacysala 954
6371 .....gctgacgacggcgccgctccacacggcggtggcg 6333
955 aspargproglualargargprocyseuargvalprocyproalagl 971
6332 tccggcgctcgtc.....accgtccgcgtca 6304
971 ntrp.....a 973
6303 gtagcccgagctcgccggcgacggcgctcgacggcggtcacgc 6254
973 rgleuenglalatrpseryincysseralathrcysgllyllegh 989
6253 ggtggacacggggtccagcagatgagctcgcttggaagggt..... 6210
990 glnarglnalvalcysargthrasnalaasnerleuylhisysgl 1006
6209 .....agtgaggaggtcgatggcgccggcggtc 6178
1006 ughlaspargproaspthvalglnvalysserleuproalacysglyg 1023
6177 ccgg.....tgtagagtcgg..... 6162
1023 lysnhisglnasnerthrvalargalaspvaltrpelyleu 1038
6161 .....tgcgctcagggcg 6147
1039 .....thprogluylgl 1043

```

```

6146 CGTCCGGGTGAACCCGTCGCGCGCTGAACACTCCGCTCC 6097
1043 ntrpvalproglinserylpro..... 1050
6096 GCACGGCCCCGGCGCATCGCGGACACCAACCATCATCACCAGCCCC 6047
1051 .....LeuHisProIleAsnLysIleSerSerMetCysAlaIaIaLupro 1065
6046 CAGGCACTACACCGGCATCCCGCTCAGACACCCCATCGGACCCACTCTCA 5997
1066 CysThrGlyAspArgSerValIlePheCysGlnMetGlyValLeuAspArgTy 1082
5996 CCAAGCTCCCGACACCCAGACCAACCCACCCACCCCATCCCGGAAA 5947
1082 rcysserileproglityrhisargleucscysvalsercysileysl 1099
5946 CGCAGCGCCCTCCGCGCAT..... 5928
1099 ysAlaserGlyProAsnProGlyProAspProGlyProThrSerLeuPro 1115
5927 .....GACGACCCCAATACCCGCGATCCCGCAACTCCCGGAAACCCACCA 5883
1116 ProPheSerThrProGlySerProLeuProGlyProGlnAspPro..... 1130
5882 CCCCACCCGACACCCCGACACCCACCCACCCGCGCGGACCCGACCG 5833
1131 ....AlaAspAlaIaIaLuproProGlyLysPro..... 1140
5832 AACTCCAGCACTCCGACACCCCGCGGAACTCCCGCAACACCGGCTCCAT 5783
1141 ..ThrGlySerGluAspHis...GlnHisGlyArgAlaThrGlnLeuPro 1155
5782 CAACACCGAATGAACCCATGCGACATCCAAACCGCGGACATCCACCC 5733
1156 GlyAlaLeuAspThrSerSerProGlyThr...GlnHisProPheAlaIpr 1171
5732 CATCCCGCATCAGCAGGACACCCGCGGAACTCCCGCAACACCGGCTCCAT 5683
1171 ogLutThrProIleProGlyAlaSerIleSerProThrThrProG 1188
5682 GACACACCCGACCGCGGCGCATTCACCGCGCAACCGACGACCACTC 5633
1188 GlyLeuProThrGlyThrGlnThrProThrProValProGlnAsp 1204
5632 CCCCACCCCTCAACACCCCGCGGACCGACTCCGACCCCGCACCG 5583
1205 LysGlyGlnProGlyLysAspLeuArgHisProGlyThrSerLeuProAl 1221
5582 ACCACATCCCCC.....CACCCACCGCGCAACCCACCCCATC 5548
1221 aaspleuProGlyArgProPro...GluProCysHisProThrGlyThr 1236
5547 AACCCACCCCGCGCACCAACCGGACCGCATCCCAACGACCA 5499

```

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-804-227C-7

seq_documentation block:

```

; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

```

alignment_scores:

Quality	Ratio	Length	Gaps
280.50	0.467	1454	80
Percent Similarity: 41.334	Percent Identity: 21.802		

alignment_block:

US-09-938-330-20 x US-08-804-227C-7/rev ..

Align seg 1/1 to reverse of: US-08-804-227C-7 from: 1 to: 44377

```

36 SerAspTyrgLysValThrValProCysSerThrAspPheArgLysArg 52
9216 TCACCCAGCGCGCTCCCGGTACGTCGCGCGCTCCACACATCC..... 9175
52 eLeuSerHisValAlaSerGlyProAlaAlaIaIaSerAlaGly.SerMet 68
9174 ....ACGTGCGGCGGGAACAACCCCGCATCAGCAGCGCTCCAGGA... 9132
69 ValValAspThrProThrGlnLeuProArgHisSerSer.HisLeuArg 85
9131 .....TCACCGCGCTCCCGCGGACCACTCCGCGCGCGGCGGACGCA 9091
85 AlaIaArgSerProLeuHisProGlyGlyThrLeuThrProGlyAlaVal 101
9090 TTGCTCCG.....ACGCTCTGATTCACGCGCATC 9062
102 GlyArgHisSerLeuTyrrPheAsnValThrVal.....PheGlyTy 115
9061 GCGCGGACGCGCGGACGACACCCGATGACGCTGCGGCGCGGCGGAGAA 9012
115 sGluLeuHisLeuArg.....LeuArgProAlaIaIaGlyLeu 128
9011 GCGGCTCCACGACGACACCCACACCCCTCCGCGCGCGCGCTCCGCTC. 8963

```


7409 7409
647 uLeuIleCysGlnSerAlaAspThrIleVal..... 659
7408CTGCTCAGGTAGCGGCGCGTGTCTCCGCGAGTACGGCG 7363
660PheMetAsnGlnValIleHisAspGly 668
7362 TCACCCGGCGGAGGTGAGCGAGTGCAGACACCGTCTGTGAGCAAC 7313
669 ThrArgCysSerTyrArgAspProTyrSerValCysAla..... 681
7312 TGCCTGCTCAGTACGCGAGCGGCGCTCTCCGGGAGCGCCGGA 7263
682ArgGlyGlnCysValProVal.....GlyCysAspGly 693
7262 GAGCGCGCGCGAGTGTCTCCGCGCGAGTGCAGCGCTGTGAGCGCGC 7213
693 IuValGlySerMetLysAlaAsp.....AspLysCysGlyValCysGly 707
7212 ACTTGGAAACGAGCGCGCTCAGCTCTGAGAGATTGCGCCACTCGGG 7163
708 GlyAspAsn.....SerHisCysArgThrValLysGlyThrLeuG1 721
7162 GCG 7119
721 LysAlaSerLysGlnAlaGlyAlaLeuLysLeuVal..... 733
7118 CCGACCGCGCGAGCTCGCGGAGCTGTGCACCCAGCGCGCGTGC 7069
734GlnIleProAlaGly.....AlaArgHisIleGln 743
7068 CCGGTGATCAGTACGTGCGCGAGGAGCAACCGGTGTCGCGGACAGCAC 7019
744 Ile.GluAlaLeuGlnLysSerProHisArgSerValLysAsnGln 760
7018 GTCCAGAGGTGTCTCTCCGCTGCTCTCCGCGCGCGGTGCGCGGCGCG 6969
760 aI.....ThrGlySerPheIleLeuAsnPro 768
6968 TGTGACGCTGCGAGCGGAGCGCGGAGCGGAGCG..... 6934
769 LysGlyLysGlnAlaThrSerArgThrPheThrAlaMetGly.....Le 783
6933 ...GGCGGGAGACACACGCGGAGCATGAGCGCGGAGCGCGTACGCG 6887
783 uGluTrp..... 785
6886 GAACGTGTTCTCTCCCGCGGAGCGAGCGCGCGAGCACTACGAGCA 6837
786GluAspAlaVal.....GluAspAlaLys..... 793
6836 CCGCGGGGAGCGACCTCCCGGGGAGTCCGCAAGCGCGCGCGAGCGCGCG 6787
794 ...GluSerLeuLysThrSerGlyProLeuProGluAlaIleAlaIle 809
6786 CCCAGTTTCAGCGCGGAGCTGTCTCCAGTCCCGACAGCTGAGCG..... 6742
809 uAlaLeuProProThrGlnGly.....GlyProArgSerSerLeuAla 824
6741CCCCGAGGAGAGGGGTCTGTCTCGGCGCGCGCGCGCGCGCT 6699
824 YLysTyrValIleHisGlnAspLeuLeuProLeuIleGlySerAsnAsn 840
6698 CACGGGTGTATCCAGAGC.....GGGCTTGGGG 6667
841 ValLeuLeuGlnGluMetAspThrTyrGluTrpAlaLeuLysSerTrp... 856
6666 ACCGCC.....TCGCGCAGCGCGCTGGAC 6644
857 ...AlaProCysSerLysAlaCysGlyGlyGly.IleGlnPheThrLys 871
6643 GAGGGCAGCGGCTCGCGCGCTCGCGCGCGCGCGCGGAGCCACAGAA 6594
872 TyrGlyCysArgArgArgArgAspHisHisMetValGlnArgHisLeu 888
6593 CTGGGGTGGTGGCGGGGAGCGCGCTCAC..... 6567
888 sAspHisLysLysArgPro..... 894
6566CGCGGGCGCGCAACAGTCCCGGCTGCTCCCGGCTGT 6527
895 ...LysProIleArgArgArgCysAsnGlnHisProCysSerGlnProVal 910
6526 ACCGACCGGACGCGCGGTGTGAGCTGTGCGCGCGCGCGCGGAGATT 6477
911 TrpValThrGlnGluTrpGlyValCysSerArgSerCysGlyLysLeuG1 927
6476 CAGTCTCATTCGGGGGGGAGTCCGAGTCCGCGGTCTCTGTCTCGGG 6427
927 yVal.....GlnThrArgGlyIleGlnCysLeuLeuP 938
6426 GCGAGAGCGACATGACCGCGAGCGCGCGGTGTCTCGCGC..... 6387
938 roLeuSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaGly 954
6386GCTGACGAGCGCGCGCTCCCAAGCGCGGTAGCGGA 6348
955 AspArgProGluAlaArgArgProCysLeuArgValProCysProAlaG1 971
6347 TCCGCGCGCTGTCTG.....ACCGTCCCGCTCA 6319
971 nTrp.....A 973
6318 GTGGCGCGAGCTCGCGCGGAGCGCGCGCTCGAGCGCGTACAGCG 6269
973 rGleuGlyAlaTrpSerGlnCysSerAlaThrCysGlyGlyIleGln 989
6268 GGTGGAGACGGGTCTCAGCGAGTACCGGTCTCGGTGGAAGGGT..... 6225
990 GlnArgGlnValValCysArgThrAsnAlaAsnSerLeuGlnHisCysG1 1006
6224AGCTGGGAGGTGATGCGCGCGCGGTCTC 6193
1006 uGlyAspArgProAspThrValGlnValCysSerLeuProAlaCysGly 1023
6192 CCGG.....TGTGAGTCTCG..... 6177
1023 LysAsnHisGlnAsnSerThrValArgAlaAspValTrpGlnLeuGly... 1038
6176TGGCTCCAGGCGG 6162
1039ThrProGlnGly 1043
6161 CGTCCCGGGTACACCGTGCAGCGCGCGCTCTGAACTCTCCGCTCC 6112
1043 nTrpValProGlnSerGlyPro..... 1050
6111 GCAGGGCGCGCGCATCTCGCGGAGCACACACATCATACCGCGCGCG 6062
1051LeuHisProIleAsnLysLysSerSerMetCysAlaIleGluPro 1065
6061 CAGGCACTACCGCGCATCCCGTTCAGACACCCATCGGAGCCACCTCA 6012
1066 CysThrGlyAspArgSerValPheCysGlnMetGlnValLeuAspArgGly 1082
6011 CCAAGCTCCCAACCGACGACGACCAACCGCGCGCGCGCGCGCGGAA 5962
1082 rCysSerLysProGlyTyrHisArgLeuCysCysValSerCysIleLysL 1099
5961 CGCACCGCTCTCCGCGCAT..... 5943
1099 ysaLysArgLysProAsnProGlyProAspProGlyProThrSerLeuPro 1115
5942GAGGACCAATATCCCGGATCCCGCAACTCCCGCGAACCACCA 5898

077

```

208 ..ArgArgGluAlaValGlnGlnIleuTrpAlaGluProAspGlyasp... 222
      |||||
      |||||
8769 AGGGCCGGCCCTCCACACCCCAACGATAGCCACCCGACCGGATATAC 8770
223 .....LeuHisAsnGluAlaPheGlyLeuGlyAspLe 233
8719 ACTGCAGAGGTCCTCCGTAGCACAATGCCCGGTAGGTGGGGGGT 8670
233 uProAsnIleu.....LeuGlyLeuValGlyAspGlnLeuGlyAspThr. 247
      |||||
      |||||
8669 CCCCAGCTGGGTGGGTGCTGCTGTGGAGATGCCGCTGACACACCG 8620
248 .....GluArgLysArgArgHisAlaLysProGlySerTyr 259
      |||||
      |||||
8619 GTCCGGGTGCGCGGACGAGTCGCGAACGATGCCGCCGCG... 8579
260 SerTleGluValLeu.....LeuValAlaAspAspSerValAlaArgPh 274
      |||||
      |||||
8578 ...CTCCACCGCTCCACGACGTCCTCCAGCAGCAACCGCTGGCGGT 8532
274 eHisGlyLysGlnHisValGlnAsnTyrValLeuThrIleuMetAsnIleV 291
      |||||
      |||||
8531 CCATGGCCCTGGCTGCGCGCGGATAT..... 8504
291 AlaAspGluIleTyrHisAspGluSerLeuGlyValHisIleAsn..... 305
8503 .....CCGGAAGAATCCGCGTGCAGACTCGCGCGCGGA 8471
306 .....IleAlaLeuValAlaGlyLeuIleMetVa 314
8470 ACCGAGAACCCGCTCCGCTACATAGCTTCCCGCGCACCCCGGCT 8421
314 IAluTyrArgIleuSerLeuSerIleu.....IleGluArg 326
      |||||
      |||||
8420 CCGGGTCGTACAGCCCTCGACGTCGCCAGCCAGCGTGGCGGAGAGGG 8371
326 LAsn.....ProSerArgSerLeuGlnValLysArgTyr 338
8370 GAGACCGTGTGCGCTCACCGGACACAGTCCACAGAGTGTCCGCGGA 8321
339 AlaHisSerGlnArgGlnAspProSerHisIleGlnHisAspHis 355
      |||||
      |||||
8320 CCGCCACAC...GCCGGGTAAAGGACACGCGCATGCCACGATCCGATCG 8274
355 sValValPheLeuThrArgGlnAspPheGlyProSerGlyTyrAlaProV 372
      |||||
      |||||
8273 GCTACGCTGTGCGGCATGACCGGGGTGGGGGGC..... 8237
372 aThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAsp 388
8236 .....CGCTGCCGGTCCGCTCCGCTGTC 8213
389 GlyPheSerSerAlaPheValIle..... 396
      |||||
      |||||
8212 GGG...ACCGAACAGTTCGTGTGTGACAGTGGAGGCGAGCGAGGGGG 8166
397 .....AlaHisGluThrG 401
8165 TCGGGGTGTCGAGAGCAGCGTGTTCGCGAGTTCAGAGCCGCGGAGCG 8116
401 LHisValLeuGlyMetGlnHisAspGlyGlnGly..AsnGlyCysAlaAs 417
      |||||
      |||||
8115 GTCCAGGGGTACGAGTTCGAGAACGTTGAGAGTGCAGAACCCGCGGA 8066
417 P...GluThrSerLeuGlySerValMetAlaPro.....LeuV 429
      |||||
      |||||
8065 CCGGAAGGCGAGGTCGCGGTGTCAGCGGTGCGCGAGCGCTGGCGAGGA 8016
429 aGlnAlaAlaPheHisArgPheHisTrpSerArgCysSerIleuGlu 445
      |||||
      |||||
8015 TCCCGGGGCTCCGAGCGCAAGCGTGAAGCAGACGACGAGCGCTTCC 7966

```

```

446 Leu.....SerArgTyrIleuProSer...TyrAspCys 455
      |||||
      |||||
7965 GCCGCGAGGGGCGCAGGAGAGCAGCGCGCGCTCGCCCGCTCTCTG 7916
455 sLeuLeuAspAspProPheAspProAlaTrpProGlnProProGluLeuP 472
      |||||
      |||||
7915 T.....CCGCTCCCGCGGTGCGGTGCGCGCGCGCTCC 7878
472 roGlyIleAsnTyrSerMetAspGlnGlnCysArgPheAspPheGlySer 488
      |||||
      |||||
7877 CTGGGTG.....AGAGCGGACAGGTCTCC 7852
489 GlyTyrGlnThrCysLeuAlaPheArgThrPheGlu...ProCysLys.. 503
      |||||
      |||||
7851 GCGGCAATGCGCGGACAGCGCGCGAGCGCGGCGGAGACGCGTGAAGAG 7802
504 .....GlnLeuTrpCysSerHisProAspAsnProTyrPheCysL 517
7801 CGACCGAGTACGGCGGTGACCGAGAGCCCGACAGAGGTCCTCCAGTCCA 7752
517 ystThrLysGlyProProLeuAspGlyThrGlnCysAlaPro..GlyL 533
      |||||
      |||||
7751 CCTCGCGCAGCAGAGGAGCTCTCGCGCGGTGCGCTGCTGCGGGA 7702
533 ystTrpCysPheLysGlyHisCysIleTyrPlySerProGlnGlnThrTyr 549
      |||||
      |||||
7701 GAGGTCCGCGCGCGCGCCACGCTGTCGAGAGAGTGCAGCGCGCGCTC 7652
550 GlnAspGlyGlyTyrPheSerTrpThrLysPheGlySerCysSerar 566
      |||||
      |||||
7651 GGGGTCCAGGGC.....CGTACCGCGCTCCGCTTCAAGACTTCCCG 7608
566 gSerCysGlyGlyGlyValArgSerArgSerArg..... 577
7607 CGTTCCGAGGCCATTCCGCGCGCGGCCACCGGCCCGCCAGCGATCGAG 7558
578 ..SerCysAsnAsnProSerProAla..... 585
7557 GGGGACCGGCTCCCTTGCGCGCGCGCGCTGCGCCAGCGCGTGCAGCGC 7508
586 ..TyrGlyLysArg.....ProCysLeuGlyProMetPheG 597
      |||||
      |||||
7507 GAGCTGTGGCGCGCGGTAGACGCGCTGCGCTGCGTCCCA..... 7466
597 uTyrGlnValCysAsnSerGlnGlnCysProGlyThrTyrGlnAspPhea 614
      |||||
      |||||
7465 ..CAGCGCAGATGAGGAGAGAGACAGCAAGAGCGTGCAGCGGAGGTCC 7417
614 rGAlaGlnGlnCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAla 630
      |||||
      |||||
7416 GCGGTGAC..... 7409
631 LysHisSerTrpValProTyrGluProAspAspAlaGlnLysCysG 647
7409 ..... 7409
647 uLeuIleCysGlnSerAlaAspThrGlyAspValVal..... 659
7408 ...CTGTCCAGGTAGCGCGCGCGGTGACCTTGTCCGCGACATAGCGG 7363
660 .....PheMetAsnGlnValValHisAspGly 668
7362 TCCACCGCGGTGAGGTGAGCGAGTGCAGACACGCTGTCGAGCAGCAC 7313
669 ThrArgCysSerTyrArgAspProTyrSerValCysAla..... 681
      |||||
      |||||
7312 TCCGCGCTGACATGCGCGCAGAGGAGCGCTCTGCGGAGCGCGGGA 7263
682 ....ArgGlyLysValProVal.....GlyCysAspLysG 693
      |||||
      |||||
7262 GAGCGCGCGAGTGTGCTCCGCTGCGCGACGTCGCGAGCGCTGACGGGC 7213
693 LuValGlySerMetLysAlaAsp.....AspLysCysGlyValCysGly 707

```

```

7212 ACTTGGAGACGAGCGCCCTGAGTCTCCAGAGTTCCGCCACTCCGG 7163
      ::|||::: ::: |||::: ::: |||
7218 G1YAspAsn.....SerHisCysArgThrValLysGlyIhrLeuG 721
      ::::: ::: |||::: ::: |||:::
7162 GCCCGCGCGCGCGCGCGCGCCACAGAGAGAGAGCGCGCGC.....GC 7119
      ::::: ::: |||::: ::: |||:::
721 YLysAlaSerLysGlnAlaGlyAlaLeuLysVal..... 723
      ::::: ::: |||::: ::: |||:::
7118 CCGACCGCGCGAGCTCGCGCGAGCTCGCGACCCAGCGAGCGAGCGGCC 7069
      ::::: ::: |||::: ::: |||:::
734 .....GlnLeuProAlaGly.....AlaArgHisIleGln 743
      ||| ||| ||| ||| ||| ||| ||| ||| |||
7068 CCGGTGATCATAGTACGCTCGCGAGAGACACCGGCTCGCGACAGCAC 7019
      ::::: ::: |||::: ::: |||:::
744 Ile.GlnAlaLeuGlnLysSerProHisArgSerValValLysAsnGln 760
      ::: |||::: ::: |||::: ::: |||:::
7018 GTCCGAGGTGTCTCGCTCGCTCTCGCGAGCGCGGCTGCGCGCGCG 6969
      ::: |||::: ::: |||::: ::: |||:::
760 a1.....ThrglySerPheIleuAsnPro 768
      ||| ||| ||| ||| ||| ||| ||| ||| |||
6968 TGTGAGACGTGCGAGCGAGCGCGGAGCGGAGCG..... 6934
      ::::: ::: |||::: ::: |||:::
769 LysGlyLysGlnAlaThrSerArgThrPheThrAlaMetGly.....Ile 783
      ::::: ::: |||::: ::: |||:::
6933 ..GGCGCGGAGACACACGCGGAGACATGAGCGCGGAGGCGCTACCGC 6887
      ::::: ::: |||::: ::: |||:::
783 uGluTrp..... 785
      ::::: ::: |||::: ::: |||:::
6886 GACTGGTCTCTCCCGCGGAGACGCGCGCGGAGCTAGCGAGCA 6837
      ::::: ::: |||::: ::: |||:::
786 .....GluAspAlaVal.....GluAspAlaLys..... 793
      ||| ||| ||| ||| ||| ||| ||| ||| |||
6836 CCGCGGCGGAGCACCTCCGCGGAGGTCGCCAGCGCGCGCGCGCGCG 6787
      ::::: ::: |||::: ::: |||:::
794 ..GlnSerLeuLysThrSerGlyProLeuProGlnAlaIleAlaIle 809
      ::::: ::: |||::: ::: |||:::
6786 CCCACTGTAGAGCGCGCGGAGCTGTCCGAGTCCCAACAGCTAGCG.. 6742
      ::::: ::: |||::: ::: |||:::
809 uAlaLeuProProthrlGlnGly.....GlyProArgSerLeuAla 824
      ::::: ::: |||::: ::: |||:::
6741 .....CCCCGAGGAGGAGGCTGTCCGCGCGGAGCGCGCGCGCT 6699
      ::::: ::: |||::: ::: |||:::
824 YLysTyrValIleHisGlnAspLeuProLeuIleGlySerAsn 840
      ::::: ::: |||::: ::: |||:::
6698 CAGGGGTGTATCCAGAGC.....GGGCTTCGGGT 6667
      ::::: ::: |||::: ::: |||:::
841 ValLeuLeuGlnGluMetAspThrTyrGluTrpAlaLeuLysSerTrp.. 856
      ::::: ::: |||::: ::: |||:::
6666 ACCGCG.....TCGCCAGCGCGCTGGAG 6644
      ::::: ::: |||::: ::: |||:::
857 ...AlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLys 871
      ::::: ::: |||::: ::: |||:::
6643 GAGGGGAGGAGGCGTGTCCGCGCTCGCGCGGCGGAGCGCGAGAA 6594
      ::::: ::: |||::: ::: |||:::
872 TyrGlyCysArgArgArgArgPheHisMetValGlnArgHisLeu 888
      ::::: ::: |||::: ::: |||:::
6593 CTGCGGTGTGTGGGCGGAGCGCTCAC..... 6567
      ::::: ::: |||::: ::: |||:::
888 sAspHisLysLysArgPro..... 894
      ::::: ::: |||::: ::: |||:::
6566 .....CGCGCGCGCGAGAGGTCGCCGCTGCCCTCCCGCGCTGT 6527
      ::::: ::: |||::: ::: |||:::
895 ..LysProLeuArgArgArgCysAsnGlnHisProCysSerGlnProVal 910
      ::::: ::: |||::: ::: |||:::
6526 ACCGACCGGAGCGGAGCTGTGAGCTGTGCGCGCGGAGCGAGGTT 6477
      ::::: ::: |||::: ::: |||:::
911 TrpValThrGluGluTrpGlyAlaCysSerArgSerCysGlyLysLeu 927
      ::::: ::: |||::: ::: |||:::
6476 CAGCTGCATCGCGGCGGAGAGTCGCGCTGTCTCTGTCGCGG 6427
      ::::: ::: |||::: ::: |||:::
927 yVal.....GlnThrArgGlyIleGlnCysLeuLeuP 938
      ::::: ::: |||::: ::: |||:::
6426 GCCAGAGACGACATGACCGCGAGCGCGCGGCTGCCG..... 6387
      ::::: ::: |||::: ::: |||:::
938 IoleuSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaGly 954
      ::::: ::: |||::: ::: |||:::
6386 .....GTGACGAGCGCGCGGCTGCCAAGCGAGCGGAGCGGA 6348
      ::::: ::: |||::: ::: |||:::
955 AspArgProGlnAlaArgArgProCysLeuArgValProCysProAlaG 971
      ::::: ::: |||::: ::: |||:::
6347 TCCCGCGCTGTG.....ACCGTCCCGCTCA 6319
      ::::: ::: |||::: ::: |||:::
971 nTrp.....A 973
      ::::: ::: |||::: ::: |||:::
6318 GTGCGCGCGAGCTCGCGCGCGCGCGGAGCTGACCGCGTACGCGC 6269
      ::::: ::: |||::: ::: |||:::
973 rGluGlyAlaTrpSerGlnCysSerAlaThrCysGlyGlnGlyIleGln 989
      ::::: ::: |||::: ::: |||:::
6268 GTGCGCGAGCGGCTCGCGAGCTAGCGGCTGTGGAAGGCT..... 6225
      ::::: ::: |||::: ::: |||:::
990 GlnArgGlnValValCysArgThrAsnAlaAsnSerLeuGlnHisCysG 1006
      ::::: ::: |||::: ::: |||:::
6224 .....AGTGGGAGGTGATGCGCGCGCGGTGCTC 6193
      ::::: ::: |||::: ::: |||:::
1006 uGlyAspArgProAspThrValGlnValCysSerLeuProAlaCysGlyG 1023
      ::::: ::: |||::: ::: |||:::
6192 CCGG.....TGTGAGTCCGCG..... 6177
      ::::: ::: |||::: ::: |||:::
1023 LysHisGlnAsnSerThrValArgAlaAspValTrpGluLeuGly... 1038
      ::::: ::: |||::: ::: |||:::
6176 .....TGGCGTCCAGGCGCG 6162
      ::::: ::: |||::: ::: |||:::
1039 .....ThrProGluGly 1043
      ::::: ::: |||::: ::: |||:::
6161 CGTCCCGGAGAACACGTCGCGAGCGCGCGCTGAAACGTCGCGCTCC 6112
      ::::: ::: |||::: ::: |||:::
1043 nTrpValProGlnSerGlyPro..... 1050
      ::::: ::: |||::: ::: |||:::
6111 GCAGGCGCGCGCGCATCGCGGACACCCACCATCATCATCACCGGCCC 6062
      ::::: ::: |||::: ::: |||:::
1051 .....LeuHisProIleAsnLysIleSerSerMetCysAlaAlaGluPro 1065
      ::::: ::: |||::: ::: |||:::
6061 CAGGCACTACCGCGCATCCCGCTGAGCAACCCGAGCGGAGCGGAGCTCA 6012
      ::::: ::: |||::: ::: |||:::
1066 CysThrGlyAspArgSerValPheCysGlnMetGlnValLeuAspArg 1082
      ::::: ::: |||::: ::: |||:::
6011 CCAACGTCCCAACACCGAGCACGACCAACCCACCCCATCCGGA 5962
      ::::: ::: |||::: ::: |||:::
1082 rCysSerIleProGlyTyrHisArgLeuCysCysValSerCysIleLys 1099
      ::::: ::: |||::: ::: |||:::
5961 CGCACCGCTCCCGCAT..... 5943
      ::::: ::: |||::: ::: |||:::
1099 yAlaSerGlyProAsnProGlyProAspProGlyProThrSerLeuPro 1115
      ::::: ::: |||::: ::: |||:::
5942 .....GAGCGACCAATACCCCGGATCCCACTCCCGGAGCGGAGCA 5898
      ::::: ::: |||::: ::: |||:::
1116 ProPheSerThrProGlySerProLeuProGlyProGlnAspPro.... 1130
      ::::: ::: |||::: ::: |||:::
5897 CCCACCGGACACACCGGACACCAACACACCGGCGGAGCGGAGCGG 5848
      ::::: ::: |||::: ::: |||:::
1131 .....AlaAspAlaAlaGluProProGlyLysPro..... 1140
      ::::: ::: |||::: ::: |||:::
5847 AACGCCAGACACTCCACACCCCGGAACTCCCGCAACCGGCTCAT 5798
      ::::: ::: |||::: ::: |||:::
1141 ..ThrGlySerGlnAspHis...GlnHisGlyArgAlaThrGlnLeuPro 1155
      ::::: ::: |||::: ::: |||:::
5797 CACACCGAATGAAACCATGCGACATCCAAACCGCGGAGCTGAC 5748
      ::::: ::: |||::: ::: |||:::
1156 GlyAlaLeuAspThrSerSerProGlyThr...GlnHisProPheAla 1171
      ::::: ::: |||::: ::: |||:::
5747 CATCCCATCAGGAGGAGCAACCGAGCTCCAGACACCATCATGACCC 5698
      ::::: ::: |||::: ::: |||:::
1171 oGluThrProIleProGlyAlaSerTrpSerIleSerProThrProG 1188
      ::::: ::: |||::: ::: |||:::
5697 GACAACACACGAGCGGCGCATTCACCGCGGAGCGGAGCGGAGCGGAG 5648
      ::::: ::: |||::: ::: |||:::

```



```

1188 1y61yLeuProTrpGlyTrpThrInThrProThrProValProGluAsp 1204
      |||
5647 CCCCAACCCCTACACACACCCCGCCACCGACGACGACGACCCCGCCACCG 5588
1205 1y5g1yGlnProGlyGluAspLeuArgHisProGlyThrSerLeuProAl 1221
      |||
5597 ACCACATCCCGCC.....CACCCACCGGACACCCACCCCATC 5563
1221 aaSPLeuProGlyArgProPro...GluProCysHisProThrGlyThr 1236
      |||
5562 AACCCACCCCGCCGACCAACCGACGACGACATCACCCACACGACACACA 5514
seq_name: /cgn2_6/plodata/2/ina/bb_COMB.seq:us-09-351-414-1

```

```

documentation_block:
  sequence 1, Application US/09351414
  Patent No. 6265199
  GENERAL INFORMATION:
  APPLICANT: Sheppard, Paul O.
  APPLICANT: Baindur, Nand
  APPLICANT: Deisher, Theresa A.
  APPLICANT: Bishop, Paul D.
  TITLE OF INVENTION: DISINTEGRIN HOMOLOG
  FILE REFERENCE: 98-29
  CURRENT APPLICATION NUMBER: US/09/351,414
  CURRENT FILING DATE: 1999-07-09
  NUMBER OF SEQ ID NOS: 13
  SOFTWARE: FastSeq for Windows Version 3.0
  SEQ ID NO 1
  LENGTH: 2268
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (3)...(2090)
  NAME/KEY: misc_feature
  LOCATION: (1)...(2268)
  OTHER INFORMATION: n = A,T,C or G
  US-09-351-414-1

```

```

alignment_scores:
  Quality: 272.00      Length: 698
  Ratio: 0.842        Gaps: 35
  Percent Similarity: 46.275  Percent Identity: 21.347

```

```

ment_block:
  09-938-330-20 x US-09-351-414-1 ..

```

```

Align seg 1/1 to: US-09-351-414-1 from: 1 to: 2268

```

```

108 pheasnValThrValPheGlyLysGluLeuHisLeuArgLeuArgProAs 124
      |||
90 TTCAGATTGAGGCTTGGGCTCCAAATTCATCTTGACCTCATCTACTGAA 139
124 naGArgLeuValValProGlySerSerValGluTrpGlnGluAspPhea 141
      |||
140 CAATGCTTGTG.....TCTCTGATTATG 165
141 rgGluLeuPheArgGlnProLeuArg.....Gln 150
      |||
166 TGGAGATTCTACTAGAAATGGAAACCAACAGTACTCTAAGGTTGAGAG 215
151 GlucysValTrpThrGlyValThrGlyMetProGlyAlaAlaValAl 167
      |||
216 CACTGTACTACATGGAAGCATCAGAGCGTCAAGACTCCAGAGTGGC 265
167 alleSerAsnCysAspGlyLeuAlaGlyLeuLeuLeuArgThrAspSerThra 184
      |||
266 TCTGTCAACTGCATGACCTTCATGCACTGTTTGAAGATGATACCTTGC 315
184 spPhePheLeuGluProLeuGluArgGlyGlnGlnGluLysGluLaser 200

```

```

316 TGTATATGATAGAGCCACTAGAGCTGTGATGATGAGAAA...AGCACA 362
      |||
201 GluArgThrHisValValTrpArgGluAlaValGlnGlnGluTrpAl 217
      |||
363 GGTGCACACCATATATATCCAGAAAACCTTGGCAGACAGTATTTT... 407
217 aGluProAspGlyAspLeuHisAsnGluAlaPheGlyLeuGlyLeuP 234
      |||
408 .....AACCAATGA 417
234 roAsnLeuLeuGlyLeuValGlyAspGln.....Leu 244
      |||
418 AGAATCTCACTATGGAAGAGGTACCAAGTGCCCTTCTCTCTGTGAATTA 467
245 GluAspThrGluArgLysArgArgHisAlaLysPro.....GlySerTy 259
      |||
468 CAGTGGTTGAAAAGAGAGAGAGAGAGTGAATTCATCAGCTGATATTT 517
259 rSer.....1leGluValLeuLeuValValAspAspSerValY 272
      |||
518 TGAAGAAAATGAATATTTGGAATTAATGATGATGTAATGACAAACGTT 567
272 alArgPheHis.....GlyLysGlnHisValGlnAsnTrpValLeuThr 286
      |||
568 ATAGAACATCGCTCTTCTCATCATCAATACACAACTTTGCAAAAGTCC 617
287 LeuMetAsn1LeuValAspGlu1LeuTrpHisAspLysSerLeuGlyAla 303
      |||
618 GTGTCACAACTTGTGATTTATTTACAGAGAGAG..... 653
303 s1LeAsn1LeuAlaLeuValArgLeu1MetValGlyTrpArgGlnSerL 320
      |||
654 .CTCAACACCAAGGTTGTCTGTGCTGTAAGACCTGAGTGAAGAG 702
320 euserLeu1leGluArgGlyAsnProSerArgSerLeuGlnGlnValCys 336
703 ATCAGATTGACATCACACCAACCTGTGCAGATGCTCCATGAGTTCTGA 752
337 ArgTrpAlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHis 353
      |||
753 AATATAC.....CGCAGAGCC.....ATTAAGCAGCATGC 781
353 sasPHisValValPheLeuThrArgGlnAspPhe.....GlyProSerG 368
      |||
782 TGATGCTGTGACCTCATCTCGCGGCTGACATTTCACTAAGAGAGCA 831
368 1yTrpAlaProValThrGlyMetCysHisProLeuArgSerCysAlaLeu 384
      |||
832 GTCTGACTTACTTTGAAGGTGTCTGTCTTCGSCAAGAGAGAGTGGTGTG 881
385 AsnHisGluAspGlyPheSerSerAlaPhe.....Val1leAlaHisG 399
      |||
882 AAT...GAGTATGCTCTCCAAATGGCAGTGGCACAAGTATATCCAGAG 928
399 uThrGlyHisValLeuGlyMetGlnHisAspGlyGlnGly..... 412
      |||
929 CTTGGCTCAAAACCTTGAATTCAAATGGCAACCTTTACCAAGAACCA 978
413 .....AsnGlyCys...AlaAspGluThrSer 420
      |||
979 AATGTGACTGCACAGAAATCTGGGCTGTGCTGCATCATGAGAGAAACAG 1028
421 LeuGlySerValMetAlaProLeuValGlnAlaAlaPheHisArgPhe 437
      |||
1029 GTGTCC.....CATTCGAAA 1045
437 sTrpSerArgCysSerLysLeuGluLeuSerArgTrpLeuPro.....S 452
      |||
1046 ATTTTCAAGTGCAGCATTTTGGAGATATGAGACTTTTTCACAGAGAGAG 1095
452 eTrpLysCysLeuLeuAspAspPro.....PheAspProAlaTrp 465
      |||

```

```

1096 GTGAGGCTGCTTTCAACAGGCCAACAACCTATTGAGCCACG... 1142
466 ProglInProglInleupProglIleasnIYrSerMetaspIglucy 482
1143 .....GAAATG 1147
482 sArgpheaSpheGlySerIYrGInIthrcysleualapheargthrp 499
1148 T.....GAAATGATAC.....G 1161
499 hegluprcysIysGInleuprcysSerHisProaspsanprotyrpe 515
1162 TGGAACTGGGAGAGGTGATGTGGTTTCATGTGGAATGCTATGA 1211
516 CysIysThrIysIysGlyProProleuaspglyThrGluCysalaprogl 532
1212 TTATGCTGTAAAGAAATGTTCCCTCCCAACGGGGCTCACTGACGACGG 1261
532 YLSTPrcyspheIysGlyHisCysIletrp..... 542
1262 GCCCTGCTGTAAACATCTCATGTCTTTTTCAGCCACGAGGGTATGAAT 1311
542 ..... 542
1312 GCCCGGATGCTGTACAGAGTGTATATCTGAATATTTCTACTGGAGAC 1361
543 .....LysSerProglInIthrcysIysGlyThrSerse 557
1362 TCTGTGATGCTCCCAACCAATCTCATAGACAGGGA..... 1400
557 rTrpThrIyspheIysSerArgSerArgSerCysasnProserProalIYr 570
1401 .....TATGATGCAATCAAAATCAGGGCCGCTGTACAAAG 1437
570 IYgIYValArgSerArgSerArgSerCysasnProserProalIYr 586
1438 GCGAGTGCAGACCAAGACACACAGTGTATGAC.....ATCTGG 1478
587 GLYgIYArgpProGlyLeuGlyPromePhegluTYrGInValCysasns 603
1479 GGAACAAAGGCTGCAGGGTCTGACAAATTTGCTATGAAAGCTGATAC 1528
603 rGluGluCysProglIthrcysIyspheArgAlaGInGlnCysAlaL 620
1529 AGAA.....GGCACTGAGAGGA.....ACTCGGGGA 1557
620 ysArgAsnSerTYrTYrValHisGlnAsnAlaIYsHis..... 632
1558 AGGATGAGAGCCGTTGAT...CAGTGCAGCAAAACATGATGTCTCTGT 1604
633 .....SerTrpValProtyrGluProaspsaspAlaG 644
605 GGATTCCTACTCTGTACCAATCTTACTGAGCTCCAGCTATGTGTCAC 1654
644 nIysCysGlnleuIle..... 649
1655 TCAGGTGAGATCATTTCCAACTCTTCTACATCAAGCCGGGTGATG 1704
650 ..CysGlnSerIle.....AspThrGlyAspValIlePhe 660
1705 ACTGAGTGTGTCCTCATGTTTACTGATGATGATGATGATGATGAT 1754
661 MetAsnGlnValIleHisAspGlyThrArgCysSerTYrArgAspPro 677
1755 .....GTAGAGATGGAAGCCCATGTGGC.....CCGTC 1783
677 rSerValCysAlaArgGlyGluCysValProValGlyCysAspIysGlu 694
1784 TATGATGTGTTAGATCGGAAGTGCCTCAAAATTCAGGCC...CTAATA 1830
694 aIGIYserMetIysAlaAspAspIysCysGlyValCysGlyGly 708
1831 TAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGC 1874

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-243-542-8
seq_documentation_block:
; Sequence 8, Application US/08243542
; Patent No. 552526
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: EXI, MITSURU
; TITLE OF INVENTION: MDC PROTEINS AND DNAs
; TITLE OF INVENTION: ENCODING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLYNN, THIEL, BOUPELL & TANIS P.C.
; STREET: 2026 Rambling Road
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49008-1699
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT Compatible
; OPERATING SYSTEM: MS-DOS 5.0
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/243,542
; FILING DATE:
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 5-136602
; FILING DATE: 14 MAY 1993
; APPLICATION NUMBER: JP 5-257455
; FILING DATE: 22 SEPTEMBER 1993
; APPLICATION NUMBER: JP 6-49904
; FILING DATE: 23 FEBRUARY 1994
; APPLICATION NUMBER: JP 6-73328
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: JP 6-84470
; FILING DATE: 22 APRIL 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Teriyence F. Chapman
; REGISTRATION NUMBER: 32 549
; REFERENCE/DOCKET NUMBER: Furuya Case 1313
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (616) 381-1156
; TELEFAX: (616) 381-5465
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; LIBRARY: human fetal brain cDNA library
; FEATURE:
; NAME/KEY: 3' UTR
; LOCATION: 2308..3183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2307
; US-08-243-542-8

alignment_scores:
Quality: 267.00 Length: 1196
Ratio: 0.560 Gaps: 66
Percent Similarity: 39.863 Percent Identity: 21.321

alignment_block:
US-09-938-330-20 x US-08-243-542-8 ..

```

Align seg 1/1 to: US-08-243-542-8 from: 1 to: 3183

```

59 GLYProAlaAlaAla.....SerAl 65
   |||||.....
76 GGTGCTGGTGGAGCTGTGGATGGGGGGCTTACCCAGCTGGAGGCC 125
   |||||.....
65 aclySerMetValValAspThrProThrLeuProArgHisSer. 81
   |||||.....
126 AGGAGCCCTGAGTTCAGCGAAGCCAGCTGTGGTTAGGAGAGCTCCG 175
   |||||.....
82 .....HisLeuArgValAlaArgSerPro 89
   |||||.....
176 GGGGAGAGTCCGAAGCAGCAGCTGACACAGAGTCCGCCAGAGACCA 225
   |||||.....
90 LeuHisProGlyGlyThrLeuThrProGlyArgValGlyArgHisSerLe 106
   |||||.....
226 .....CCAGGGGGGGGG.....CTGTCCATGTGGCCCGAGTGAAGT.. 261
   |||||.....
106 uThrPheAsnValThrValPheGlyLysGluLeuHisLeuArgLeuArgp 123
   |||||.....
262 .....TTGCTCATCCAGCCTTCAACTCACTCACTCACTGGAGCTGGAGC 307
   |||||.....
123 roAsnArgLeuValValProGlySerSerValGluThrGlnGlnAsp 139
   |||||.....
308 TGAACCCACCTCTCTC.....TCCTGCAATAGCTGGAG... 342
   |||||.....
140 PheArgGluLeuPheArgGlnProLeuArgGln..... 150
   |||||.....
343 ...GCCCACTTACGCCGGAGGAGGACACCCAGCAGCAGCGGGCTGG 389
   |||||.....
151 ...GluCysValThrThrGlyGlyValThrGlyMetProGlyAlaAlaLay 166
   |||||.....
390 AGACACAGCTACTACCGAGGAGACTCCGGGGAGCCCGACACTCTTCG 439
   |||||.....
166 aAlaIleSerAsnCysAspGlyLeuAlaGlyLeuIleArgThrAspSer 182
   |||||.....
440 CCGCCCTTCCACCTGCCAGGGGCTCATGGGCTTCTGTGATGGAGAC 489
   |||||.....
183 ThrAspPhePheIleGluProLeuGluArgGlyGlnGlnGlnGlnGln 199
   |||||.....
490 TTGACTTACATCGTGAAGCCCAAGAG..... 516
   |||||.....
199 aSerGlyArgThrHisValValThrArgArgGluAlaValGlnGlnGln 216
   |||||.....
517 .....GTGGCTGAGACTT 529
   |||||.....
216 rPalGluProAspGlyAspLeuHisAsnGluAlaPheGlyLeuGlyAsp 232
   |||||.....
530 GGGGAGGCCCTCAG.....GGAGCC 549
   |||||.....
233 LeuProAsnLeuLeu.....GlyLeuValGlyAspGlnLeuGlyAs 246
   |||||.....
550 CTTCGCCACCTCATTTACCGAGCCCTCTCTCCAGATCCCTCGGAGG 599
   |||||.....
246 pThrGlu..... 248
   |||||.....
600 CAGGGAACCAAGCTGTTGTGCTGTGCTGCCAGCTCGCTCTCCAA 649
   |||||.....
249 .....ArgLysArgArgHisAlaLysProGlySerTyrSer 260
   |||||.....
650 ACCGGCCGAGGCTGAGAGAGAAAGCAGGTCGGCGGGCCACCTTACA 699
   |||||.....
261 IleGlu.....ValLeuLeuValValAspAspSerVal 271
   |||||.....
700 GTGCACAGTGAACCAAGTATGTGAGCTAATGTGATCAAGACACCA 749
   |||||.....
271 lValArgPheHisGlyLysGlnHisVal.....GlnAsnTyrValL 285
   |||||.....
750 GCTTTTGACAGATGCGAGCTGCTGTCTCCACAGCAACTTTCGCA 799
   |||||.....
285 euThrLeuMetAsnIleValAspGluIleTyrHisAspGlnSerLeuGly 301
   |||||.....

```

```

800 ACTCCGTGTGAACCTGGCCGATGTATATACAGAGACAG..... 840
   |||||.....
302 ValHisIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrArgG 318
   |||||.....
841 .....CTCAACACTCGCATGTCTCTGTCTCCATGGCAACATGGGACA 884
   |||||.....
318 nSerLeuSerLeuIleGluArgGlyAsnProSerArgSerLeuGlnL 335
   |||||.....
885 TGGGAGCAAGATCCAGGCTGCGAGATGACCTCTTGAGAGACCTGGCCGGC 934
   |||||.....
335 aLysArgTyrPalHisSerGlnArgGlnAspProSerHisAlaGlu 351
   |||||.....
935 TCATGTCTTAC...CGAGGAGAGGCTGTGCTGAGCCAGTAATGCCACC 981
   |||||.....
352 HisHisAspHisValValPheLeuThrArgGln...AspPheGlyProse 367
   |||||.....
982 CAC.....CTCTTCTCGGGCAGGACTTTCGAGAGCAGAGACAG 1019
   |||||.....
367 rGlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaL 384
   |||||.....
1020 CGGGGACACCTTACGTTGGGGGCATA.....TGTCTCC 1051
   |||||.....
384 euAsnHisGlnAspGlyPheSerSer.....Ala 393
   |||||.....
1052 TGTCCCATGGCGGGGCTGTGAACGATACGGCAACATGGGGCGATGGCC 1101
   |||||.....
394 PheValIleAlaHisGlnThrGlyHisValLeuGlyMet.....G 407
   |||||.....
1102 GTGACCTTGGCCACAGCCTGGGACAGACCTGGGCTGATGTGAACAA 1151
   |||||.....
407 uHisAspGlyGlnGlyAsnGlyCys...AlaAspGluThrSerLeuGly 423
   |||||.....
1152 ACACCGGAGCTCGGCGAGGAGCTCAGATGCCAGACATGTGCTGGCT 1201
   |||||.....
423 eValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTyrSer 439
   |||||.....
1202 GCATCATGAGAGACATGGGTTTACTGTGCCCGCAAGTTTC.....T 1245
   |||||.....
440 ArgCysSerLysLeuGluLeuSerArgTyrLeuPro.....SerTyrAs 454
   |||||.....
1246 CGGTGCGACATCGAGAGATACACCACTTCTGACAGAGGTTGTGGCAG 1295
   |||||.....
454 pCysLeuLeuAspAspProPheAspProAlaTyrProGlnGly 471
   |||||.....
1296 CTGCTCTTCAACAAAGCCCTCAAGCTCTCG.....GACCCCGCAGAT 1339
   |||||.....
471 euProGlyIleAsnTyrSerMetAspGlnGlnCysArgPheAspPheGly 487
   |||||.....
1340 GCGGGAACGGCTTGTGTGAGCGAGGAGGAGTGC.....GACTGCGGC 1383
   |||||.....
488 SerGlyTyrGlnThrCysLeuAlaPheArgThrPheGluProCysLysG 504
   |||||.....
1384 TCGGTGCAAGAG..... 1395
   |||||.....
504 nLeuTyrCysSerHisProAspAsnProTyrPheCysLysThrLysG 521
   |||||.....
1396 .....TGCAGCCGC..... 1404
   |||||.....
521 lYrProProLeuAspGlyThrGluCysAlaProGlyLysTyrCysPheLys 537
   |||||.....
1405 .....GCAAGTGGCAACTGCTGCAAGAAA 1428
   |||||.....
538 GlyHisCysIleTyrLysSerProGluGlnTyrGlyGlnAspGlyG 554
   |||||.....
1429 .....TGC.....ACCTGACTCAGAGAGCCAT 1451
   |||||.....
554 yTyrPserSerTyrThrLysPheGlySerCysSerArgSerCys..... 568
   |||||.....
1452 GTGCACAGAC.....GGCTGTGCTGTGCGCGCTGCAAGTACG 1489
   |||||.....
569 .....GlyGlyGlyValArgSerArgSerArgSerCysAsnAsnPro 582
   |||||.....
1490 AACCAAGGGGTGTCTCTGCGAGAGGCCGTGAACAGTGCACATC... 1536
   |||||.....

```

```

583 SerProAlaTyrGlyArgProCysLeuGlyProMetPheGluTyrGlu 599
      ::::: ||| |||
1537 .....GCGAGACCTGCACCGGG..... 1554
599 nValCysAsnSerGluGlyCysProGlyThrTyrGluAspPheArgAlaG 616
      ::::: ||| |||
1555 .....GACTCTAGCCAGTGCCTTACCTGCAC..... 1584
616 InGlnCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHis 632
      ||| ::::: ||| |||
1585 .....AACTGGACGGCTTACTACTGT..... 1605
633 SerTrpValProTyrGluProAspAspAlaGlnLysCysGluLeuI 649
      ||| ::::: |||
1606 .....GACCATGACGACGGCGCTGTAC..... 1629
649 eCysGlnSerAlaAspThrGlyAspValValPheMetAsnGlnValValH 666
1629 ..... 1629
666 IsAspGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArg 682
      ||| ||| ||| |||
1630 .....GGAGTGCCTGCACAAACCGGACCGGAG..... 1659
683 GlyLeuCysValProValGlyCysAspLysGluValGlySerMetLysAl 699
      ||| ||| |||
1660 .....TGCCAGGTCTTGGGGCCATCGCGCTGC 1688
699 ASP.....AspLysCysGlyValCysGlyLysAspAsnSerH 712
      ||| ||| ||| |||
1689 TGATCGCTTCTGCTACGAGAGCTGAATGTGGAGGGACGGAGCTGGGA 1738
712 IsCysArgThrValLysGlyThrLeuGlyLysAlaSerLysGln..... 726
      ||| ||| ||| |||
1739 GCTGTGGCGCAGAGGATCCGCTGGCTGCAGTGCATTAACGACGAGCTG 1788
727 ..AlaGlyAlaLeuLysLeuValGlnIleProAlaGlyAlaArgHisI 742
      ||| ||| ||| |||
1789 CTGTGTGCTTCTCTCTGTGTCAACATCTCTGTGAGCTCTCTCGG..... 1833
742 eGlnIleGlnAlaLeuGlnLysSerProHisArgSerValValLysAsnG 759
1834 .....CTAGGGG 1840
759 InValThrGlySerPheIleLeuAsnProLysGlyLysGlnAlaIleHse 775
      ::::: ||| ||| ||| |||
1841 AACTGTGGGAGACATCAGTAGTGTACCTTCTACACGACGAGGCAAGAG 1890
775 rArgThrPheThrAlaMetGlyLeuGluTrpGluAspAlaValGlnAspA 792
      ||| ||| ||| |||
1891 CTGGACTGCAGGGGAGGCCACGCTGCAGCTGGCGGAGCGCTGTGACCTAG 1940
792 IaLysGlnSerLeu..... 796
      ||| |||
1941 CTATGTGAGGATGGCAGACGCTGGGGCTTAACATGTGTGCTGCAGACC 1990
797 .....LysThrSerGlyProLeuProGlnAla 805
1991 ATNCGCTGCCTGCACCTTCTGCTTCAACCTTCAACCTGCCTGCCTGC 2037
805 AlIleAlaIleAlaLeuProProThrGluGlyLysProArgSerSerL 822
      ::::: ||| ||| |||
2038 AGTGGGAGCGCGGATTTGCTCCACACGAGGG.....TCTG 2075
822 eValaTyrLysTyrValIleHisGlnAspLeuLeuProLeuIleGlySer 838
      ||| ||| ||| |||
2076 CAGCAATGAG.....GGAGT 2092
839 AsnAsnValLeuLeuGlnGluMetAspThrTyrGluTrpAlaLeuLysSe 855
      ::::: ||| ||| ||| |||
2093 GCATCTGTC.....AGCCAGACTGACAGACGCAAAAGACTGCAGTA...TC 2133

855 rTrpAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysT 872
      ||| ||| ||| |||
2134 CATAAACCCCTGCCACGCTCCACACCGGAG..... 2167
872 YrGlyCysArgArgArgArgAspHisHisMetValGlnArgHisLeuCy 888
      ||| ||| ||| |||
2168 .....AGACGAGAGAT...ATTAAGTCTCCAGCGGACACCAACA 2203
888 sAspHisLysLysArgProLysProIleArgArgArgCysAsnGlnHisP 905
      ||| |||
2204 TCATCAT..... 2210
905 rCysSerGlnProValTrpValThrGluTrpGly.....AlaCys 919
      ||| ||| ||| |||
2211 .....TGCTTC...CATCGTGGGGCTGTCCTGTGTC 2240
920 SerArgSerCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLe 936
      ||| ||| ||| |||
2241 AGCCATCGTCTGGCGGACGCGCTG..... 2267
936 uLeuProLeuSerAsnGlyThrHisLysValMetProAlaLysAlaCysA 953
      ||| ||| ||| |||
2268 .....GGATTTAAACATTCGCCGAGGAAGTCCGGA 2301
953 IaGlyAspArgProGlnAlaArgArgProCysLeuArgValProCysPr 969
      ||| ||| |||
2302 GGGGCTTAAGTGCAACCTTC..... 2322
969 oAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyG 986
      ||| ||| ||| |||
2323 .....TCCCTCAAGCTGGACCCACCTCT..... 2349
986 InGlyIleGlnIleArgGlnValValCysArgThrAsnAlaAsnSerLeu 1002
2349 ..... 2349
1003 GlnHisCysGluGlyAspArgProAspThrValGlnValCysSerLeuPr 1019
      ||| ||| ||| |||
2350 .....CGGCCCTGAACCCAGAG.....CTGGC 2372
1019 o.....AlaCysGlyLysAsnHisGlnAsnSerThrValArgAlaAspY 1034
      ||| ||| ||| |||
2373 CCCATCCAGCCACGAGGAGGACCATGCAATGTCTTCCAGTCCAAA 2422
1034 AlTrpGluLeuGlyThrProGluGly..... 1042
      ::::: ||| ||| ||| |||
2423 CCTTCAACCTCGCTCCGACAGGGTTGGGAGGGCTGTGGCCCTGC 2472
1043 GlnTrpValProGlnSerGlyPro..... 1050
      ||| ||| ||| |||
2473 CPTTGGCACCCACGAGGTGACACGAGCTGAGAGGCACCTCTCCACAGT 2522
1051 LeuHisProIleAsnLysIleSerSerMetCysAlaIleGluProCysT 1067
      ||| ||| ||| |||
2523 CCCCCACCCA.....CTCTGGGGCTCAGCCTTGA 2554
1067 InGlyAspArgSerValPheCysGlnMetGlnValLeuAspArgTyrCys 1083
      ||| ||| ||| |||
2555 CACCAACCTGCCCGCTGGAATGTAGCTTCCACCTCATGATGCCACAC 2604
1084 SerIleProGlyThrHisArgLeuCysCysValSerCysIleLysLysAl 1100
      ||| ||| ||| |||
2605 TCAACTCGGGGACCTGAGAGGATCCCCACGAGCC.....AC 2645
1100 AserGlyProAsn.....ProGlyPro...AspP 1109
      ||| ||| ||| |||
2646 CAGTGAACCTAAGCTGATGGCCCTCTTGCACACGAGCAGTGTAGAGAC 2695
1109 rGlyProThrSerLeuProProPheSerThr 1119
      ||| ||| ||| |||
2696 AGGCTCTTATCTCTCTGAGACTAGGGGAGCG 2727

seq_name: /cgn2_6/ptodata/2/ina/5A_COMP.seq:US-08-477-407-8

```

seq documentation block:

Sequence 8, Application US/08477/407
 Patent No. 5631351
 GENERAL INFORMATION:
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: EMI, MITSURU
 TITLE OF INVENTION: MDC PROTEINS AND DNAs
 TITLE OF INVENTION: ENCODING THE SAME
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.
 STREET: 2026 Rambling Road
 CITY: Kalamazoo
 STATE: Michigan
 COUNTRY: USA
 ZIP: 49008-1699
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inches, 1.44 MB storage
 COMPUTER: IBM PC/XT/AT Compatible
 OPERATING SYSTEM: MS-DOS 5.0
 SOFTWARE: WordPerfect 5.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,407
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/243,542
 FILING DATE: 13-MAY-1994
 APPLICATION NUMBER: JP 5-136602
 FILING DATE: 14 MAY 1993
 APPLICATION NUMBER: JP 5-257455
 FILING DATE: 22 SEPTEMBER 1993
 APPLICATION NUMBER: JP 6-49904
 FILING DATE: 23 FEBRUARY 1994
 APPLICATION NUMBER: JP 6-73328
 FILING DATE: 12 APRIL 1994
 APPLICATION NUMBER: JP 6-84470
 FILING DATE: 22 APRIL 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Terrence F. Chapman
 REGISTRATION NUMBER: 32 549
 REFERENCE/DOCKET NUMBER: Futuya Case 1313
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (616) 381-1156
 TELEFAX: (616) 381-5465
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3183 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 LIBRARY: human fetal brain cDNA library
 FEATURE:
 NAME/KEY: 3' UTR
 LOCATION: 2308..3183
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2307
 US-08-477-407-8

alignment_scores:

Quality: 267.00 Length: 1196
 Ratio: 0.560 Gaps: 66
 Percent Similarity: 39.883 Percent Identity: 21.321

alignment_block:

US-09-938-330-20 x US-08-477-407-8 ..

Align seg 1/1 to: US-08-477-407-8 from: 1 to: 3183

```

59 GLYPROLAlaLaLa.....SerAl 65
   |||||||.....
76 GGCTGTGTGAGACTCTGCATGGGGGGCTTACCCAGTCGGAGGCC 125
   |||||||.....
65 aclysermetValAlaAspThrProThrLeuProAlaHisSer 81
   |||||.....
126 AGGAGCCCCGAGGTGACGAGAACCCAGCCGTCTGTTGGAGAGACTCCG 175
   |||||||.....
82 .....HisLeuArgValAlaArgSerPro 89
   |||||.....
176 GGGGAGAGGTCCGAAAGCAGCAGCTGACAGAGGCTCCCGACAGACCA 225
   |||||||.....
90 LeuHisProGlyGlyThrLeuThrProGlyArgValGlyArgHisSerLe 106
   |||||||.....
226 .....CCAGGGGGGCCG.....CTGTCCATCTGGCCGAGTCACT.. 261
   |||||.....
106 uTyrPheAsnValThrValPheGlyLysGlyLeuHisLeuArgLeuArgp 123
   |||||.....
262 .....TTCCGATCCCGACGCTTCAACTCAACTCACTGACCTGGAGC 307
   |||||||.....
123 roAsnArgArgLeuValValProGlySerSerValGluTyrPGLuAsp 139
   |||||||.....
308 TGAACCCACACCTCTCTC.....TCTCGCAATACGTGGAG... 342
   |||||||.....
140 PheArgGlyLeuPheArgGlnProLeuArgGln..... 150
   |||||||.....
343 ...GCCCACTTCAGCGCGGAGGAGCAACCCAGCAGCAGCGGGCTCG 389
   |||||||.....
151 ...GluCysValTyrThrGlyValThrGlyMetProGlyAlaAlaIav 166
   |||||.....
390 AGACCACTGCTACTACAGGAGGAGCTCCGGGAGACCCGACCTCTTCG 439
   |||||||.....
166 aAlaIleSerAsnGlyAspGlyLeuAlaGlyLeuIleArgThrAspSer 182
   |||||||.....
440 CGCCCTCTCCACCTGCGCCAGGGGCTGCATGGGCTCTCTGTGATGGCAG 489
   |||||||.....
183 ThrAspPhePheIleGluProLeuGluArgGlyGlnGlnIuysGluAl 199
   |||||||.....
490 TTGACTTACATCGTGAAGCCCAAG..... 516
   |||||.....
199 aserGlyArgThrHisValValTyrArgArgGluAlaValGlnGlnGlu 216
   |||||.....
517 .....GTGCTGGACCTT 529
   |||||.....
216 rPalagluProAspGlyAspLeuHisAsnGluAlaPheGlyLeuGlyAsp 232
   |||||.....
530 GGGAGGCCCTCAG.....GGACC 549
   |||||.....
233 LeuProAsnLeuLeu.....GlyLeuValGlyAspGlnLeuGlyAs 246
   |||||||.....
550 CTTCGCCACCTCATTTACCGGACCCCTCTCTCCAGATCCCTCGGANG 599
   |||||.....
246 pThrGlu..... 248
   |||||.....
600 CAGGGAACAGGCTGCTGTTGCTGTGCTGCCAGTGGCTCTCTCAA 649
   |||||.....
249 .....ArgLysArgArgHisAlaLysProGlySerTyrSer 260
   |||||||.....
650 ACCGCGCGAGGCTGAGAAAGGAGGAGTCCGCGGGGCCACCTTACA 699
   |||||||.....
261 IleGlu.....ValLeuLeuValValAspAspSerVa 271
   |||||||.....
700 GTGCACAGTAAACCAAGTATGTGAGCTAATTGTGATCAGACACCA 749
   |||||||.....
271 lValArgPheHisGlyLysGluHisVal.....GlnAsnTyrValL 285
   |||||.....
750 GCTGTGTGACAGATGAGACAGTGGTGTCTCTCAGCAGCAACTTGGCA 799
   |||||||.....
285 eutHrLeuMetAsnIleValAspGluIleTyrHisAspGluSerLeuGly 301
   |||||||.....

```

```

800 AGTCCGTGTGAACCTGGCCGATGTATATACAGGAGCAG..... 840
302 ValHisIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrArgI 318
841 .....CTCAACACTGCATCGTCTGGTTCATGGAACAATGGGCACA 884
318 nSerLeuSerLeuIleGluArgGlyAsnProSerArgSerLeuGluGln 335
885 TGGGACAAAGATCCAGGATGACCTCGAGACCTGGCCCGCC 934
335 aLysAlaGlyThrAlaHisSerGlnGlnArgGlnAspProSerHisAlaGlu 351
935 TCATGTGTAC...CGACGGAGGGTCTGCTGAGCCCACTATGCGCAC 981
352 HisHisAspHisValAlaPheLeuThrArgGln...AspPheGlyProSe 367
982 CAC.....CTTCTTCCGGGACGACCTTCACAGACGACGACG 1019
367 rGlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaL 384
1020 CGGGGACACCTACCTGGGGGACATA.....TGCTCC 1051
384 euAsnHisGluAspGlyPheSerSer.....Ala 393
1052 TGTCCATGCGGGGGGTGTGACAGATACGGCAACATGGGGCGGATGGCC 1101
394 PheValIleAlaHisGlyThrGlyHisValLeuGlyMet.....G 407
1102 GTGACACCTTGGCCACGACCTGGGACAGACCTGGGACATGATGTGAACA 1151
407 uHisAspGlyGlnGlyAsnGlyCys...AlaAspGlyThrSerLeuGly 423
1152 ACACCGGACCTGGGACGGGACTCCAACTGTCCAGACATCTGCTGGCT 1201
423 eValMetAlaProLeuValGlnIleAlaPheHisArgPheHisTrpSer 439
1202 GCATCATGTGAGGACACTGGGTTTACCTGCCCGCCAAAGTTC.....T 1245
440 ArgCysSerLeuGlyLeuSerArgTyrLeuPro.....SerTyrAs 454
1246 CCGTGCAGCATCGACAGTACACCACTTCTCGAGAGGGTGTGTGCAG 1295
454 pLysIleLeuAspAspProPheAspProAlaTrpProGlnProGluL 471
1296 CTGCGCTTTCACAAAGCCCTCAAGCTCTG.....GACCCCCCAAGT 1339
471 euProGlyIleAsnTyrSerMetAspGlnGlyCysArgPheAspPheGly 487
1340 GGGGGAACGGCTTGTGGAGGACGGGAGAGTGC.....GACTGCGGC 1383
488 SerGlyTyrGlnThrCysLeuAlaPheArgThrPheGluProCysLysG 504
1384 TCGGTGCAGAG..... 1395
504 nLeuTyrCysSerHisProAspAsnProTyrPheCysLysThrLysG 521
1396 .....TGCACCGCC..... 1404
521 LyrProProLeuAspGlyThrGlyCysAlaProGlyLysTrpCysPheLys 537
1405 .....GCAGGTGGCAACTGCTGCAGAGAA 1428
538 GlyHisCysIleThrLysSerProGluGlnIleThrGlyGlnAspGlyL 554
1429 .....TGC.....ACCTGACTCAGAGCCCAT 1451
554 LyrSerSerTrpThrLysPheGlySerCysSerArgSerCys..... 568
1452 GTGCACGAC.....GGGCTGTGTGTGCGCGCTGCAAGTAGC 1489
569 .....GlyGlyGlyValArgSerArgSerArgSerCysAsnAsnPro 582
1490 AACCAAGGGGTGTGTCTGTCGAGAGGCCGTGAACGAGTGCAGACATC... 1536
583 SerProAlaTyrGlyLysArgProCysLeuGlyProMetPheGluTyrG 599
1537 .....GGGAGACCTGCACCGGG..... 1554
599 nValCysAsnSerGluGlyCysProGlyThrTyrGlnAspPheArgAlaG 616
1555 .....GACTTACGCACAGTGGCCCGCTCAACCTGCAC..... 1584
616 GlnGlyCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHis 632
1585 .....AAGCTGACGGTTACTACTGT..... 1605
633 SerThrValProTyrGlnProAspAspAlaGlnLysCysGluMetI 649
1606 .....GACCATGACAGAGCGCGCTGCTAC..... 1629
649 eCysGlnSerAlaAspThrGlyAspValAlaPheMetAsnGlnValL 666
1629 ..... 1629
666 LysAspGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArg 682
1630 .....GGAGTGGCTGCACAAACCCGGGACCGCAG..... 1659
683 GlyGlyCysValProValGlyCysAspLysGluValGlySerMetLysAl 699
1660 .....TGCCAGGTCTTTGGGGCCATGCGGCTGC 1688
699 aAsp.....AspLysCysGlyValCysGlyGlyAspAsnSerH 712
1689 TGATGCGTTCCTGCTACGAGAACCTGATGTGAGGGACGACCGAGCGGGA 1738
712 LysCysArgThrValLysGlyThrLeuGlyLysAlaSerLysGln..... 726
1739 GCTGTGGGCGCAAGGATCCGGCTGGGTCCAGTGCAGTTAAGCAGACGCTG 1788
727 ..AlaGlyAlaLeuLysLeuValGlnIleProAlaGlyAlaArgHisI 742
1789 CTGTGTGCTCTTCCTCCTGTGTGTCAACATCTGTGAGACTCTCTCG..... 1833
742 eGlnIleGluAlaLeuGlnLysSerProHisArgSerValValLysAsnG 759
1834 .....CTAGGGG 1840
759 LysValThrGlySerPheIleLeuAsnProLysGlyLysGluAlaThrSe 775
1841 ACCTGTGGAGACATCAGTAGTGTCACTTTCACACGAGGCAAGAG 1890
775 rArgThrPheThrAlaMetGlyLeuGlyLysGlyLysAspAlaValGluAsp 792
1891 CTGACCTCAGGGGAGGAGCCAGCTGACCTGGCGACGCTGTGACCTGAG 1940
792 LysGluSerLeu..... 796
1941 CTATGTGAGAGATGGCACAGCTCGGGCTAACATGTTGCTGCTGAGAC 1990
797 .....LysThrSerGlyProLeuProGluAl 805
1991 ATCGCTGCGTGCACCTTCTGCTTCAACTTCACACCTGCGCGG...GC 2037
805 aLleAlaIleLeuAlaLeuProThrThrGlnGlyGlyProArgSerSerL 822
2038 AGTGGGAGCGCGGATTTGCTCCACACCGCGG.....TCTG 2075
822 euAlaTyrLysTyrValIleHisGluAspLeuProLeuIleGlySer 838
2076 CAGCAATGAG.....GGAAGT 2092
839 AsnAsnValLeuLeuGluGlnMetAspThrTyrGlnThrAlaLeuLysSe 855
2093 GCATCTGTC.....AGCCAGACTGACAGCAAAAGACTGCAGTA...TC 2133

```

```
855 rTTPAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysT 872
      |:::|||||
2134 CATACCCCTGCCCAAGTCCACCACGGGG..... 2167
      |||
872 YrGlyCysAlaGArgArgArgPheHisMetVal.GlnArgHisLeuCy 888
      |||||||||
2168 .....AGACGAGAGAT...ATAAGGTCCACGCGCACCAACA 2203
      |||
888 sAsPHisLysLysArgProLysProIleArgArgCysAsnGlnHisP 905
      |||
2204 TCATCAT..... 2210
905 rOcYsSerGlnProValTTPValThrGluLutPgly.....AlaCys 919
      ||:::|||||
2211 .....TGGCTC...CATCGTGGGGCTGTCTCTGTTGC 2240
920 SerArgSerCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLe 936
      ||:::|||||
2241 AGCATCTCTCTGGCGCACGGCTG..... 2267
936 uLeuProLeuSerAsnGlyThrHisLysValMetProAlaLysAlaCysA 953
      |||
2268 .....GGGATTTAAACATTCGCCGAGAGAGTCCGGA 2301
953 laGlyAspArg.ProGlnAlaArgArgProCysLeuArgValProCysPr 969
      |||
2302 GGGGCTTAAGTGCACCCCTTC..... 2322
969 oAlaGlnTPArgLeuGlyAlaTPSerGlnCysSerAlaThrCysGlyG 986
      |||
2323 .....TCCCTCCAAAGCTGGCACCCACCGCTT..... 2349
986 lUGlyIleGlnGlnArgIlnValCysArgThrAsnAlaAsnSerLeu 1002
2349 ..... 2349
1003 gLYHisCysGluGlyAspArgProAspThrValGlnValCysSerLeuPr 1019
      |||||
2350 .....CGGCTTGAACCAAGG.....CTGCC 2372
1019 o.....AlaCysGlyGlyAsnHisGlnAsnSerThrValAlaArgAlaAspV 1034
      |:::|||||
2373 CCCATCCAGCCACGAGGAGGAGGACCATGCAATGTCTTCCAGGTCCAAA 2422
1034 aLTPGluLeuGlyThrProGluGly..... 1042
      ::::|
2423 CCTTCAACTCTGCTGCTCCGAGGGGTTGGGTGGGGCTGTGGCCTGC 2472
1043 GlnTPValProGlnSerGlyPro..... 1050
      |||
2473 CCTTGGCACCAACGAGGTGGACCAAGCCTGGAGGGCACTTCTCCACAGT 2522
1051 .LeuHisProIleAsnLysIleSerSerMetCysAlaAlaGluProCysT 1067
      |||||
2523 CCCCACCCA.....CCTCTCGCGGCTCAGCCTTGCA 2554
1067 hnGlyAspArgSerValPheCysGlnMetGlnValLeuAspArgTyrCys 1083
      |||
2555 CACCCACTGCCCGTGTGAATGTAGCTTCCACCTCATGATGCCACAGC 2604
1084 SerIleProGlyTyrHisArgLeuCysCysValSerCysIleLysLysAl 1100
      |||
2605 TCAACTGGGGGACACTGGAGGATGCCCCAGACAGCC.....AC 2645
1100 aSerGlyProAsn.....ProGlyPro...Asp 1109
      ::|||
2646 CAGTGGACCTTAGCTGTGATGGCCCTCTTGCACACAGGAGGCTGAGACC 2695
1109 rOGlyProThrSerLeuProProPheSerThr 1119
      |||
2696 AGGCTTTATCTCTGTGGACCTAGGGGAGC 2727
```

